

No easy answers

Britain's regulator has taken a sensible approach to the fraught question of what kinds of genetic testing should be permitted on embryos.

There is no gene for the human spirit, declared the producers of the 1997 movie *Gattaca*, which related the sobering tale of a eugenic community that had overestimated the usefulness of genes in predicting a person's destiny. The movie's hero, Vincent, overcomes a damning genetic prognosis and resultant discrimination to become an astronaut, while the fall guy, Jerome, fails to achieve the athletic prowess supposedly encoded in his DNA.

Gattaca echoed widespread public disquiet about the possible use and abuse of genetics in the selection of embryos that are free of particular disease-causing genes. Opinions on the matter range from those who think the market should be allowed to determine the use of such testing, as occurs in the United States, to others who would ban embryo selection altogether.

In nations that choose to regulate genetic tests, the real question is how far testing should be allowed to go. They are already used for the genetic mutations that cause cystic fibrosis, where a severely adverse outcome for the child is reasonably certain, and quite early in life. The question now arises of what to do regarding genes (such as *BRCA1*, which leads to breast cancer) where the outcome may not be fatal and comes later. Decisions will soon have to be made regarding tests for genes whose implications are even less specific.

How should societies wrestling with this question proceed? Different nations have sharply different regulatory arrangements, often heavily influenced by their histories of genetics and eugenics. In Germany, for example, an Embryo Protection Law currently prohibits the genetic testing of embryos altogether.

In the United States, by contrast, memories of forced sterilization when eugenics was in fashion early in the twentieth century led to a series of Supreme Court judgements that support the individual's right to procreate. Despite the concerns of some religious groups, this has made the federal government reluctant to restrict such freedom. The result has been a free-for-all that doesn't even offer anxious would-be parents adequate means of assessing the scientific merit of the tests to which they are subjected.

The middle ground is being sought in Britain, where the Human Fertilisation and Embryology Authority (HFEA) regulates both

fertility clinics and embryo research. At a meeting in Belfast last week, the authority ruled that the testing of embryos should be permitted, in certain circumstances, for genes that are not certain to cause severe illness, but merely likely to do so.

The HFEA has a reasonable track record of consulting openly with people of many different backgrounds and views. Its job is to interpret the law in the light of technological developments and take the necessary decisions about what should or should not be permitted. The authority's 18 members represent a mix of scientists, ethicists, law experts, theologians and members of the public. It strives to be independent, and neither its chair nor its deputy chair may be scientists or clinicians involved in fertility treatment.

Last week's decision followed a fact-gathering exercise, publication of a consultation document, a public meeting held last year in London, and consideration of the issue by the authority's ethics and law committee. The HFEA already allowed genetic tests to be used for conditions where a single gene is certain to lead to disease; the new decision extends this to conditions where the gene is only likely to cause disease, and allows affected families and their clinicians to have a say in how severe they consider the disease to be. For the time being, clinics will have to apply to a licence committee for permission to treat each family that approaches them.

The decision could not have pleased everyone, and critics voiced strenuous objections to what they regard as a 'slippery slope' towards the more permissive use of genetic testing. But the decision to allow individual families and their doctors some discretion in determining just what constitutes a serious condition is preferable to the alternative of decreeing prescribed lists of tests that are banned or permitted; the latter path would fall far closer to state-mandated eugenics. The authority should be applauded for finding a middle road that will allow this potentially valuable technology to move forwards, under careful supervision. ■

"The HFEA ruled that the testing of embryos should be permitted for genes that are not certain to cause disease, but are merely likely to do so."

Infection biology

Immunology and microbiology come together to fight disease.

It is hardly surprising that scientific papers on human pathogens are often peppered with military analogies. Traditionally, the study of infectious disease has been considered as a conflict in which pathogens are the enemy, to be sought out and destroyed. The

success of antibiotics in the twentieth century reinforced the notion that infectious diseases could be defeated, if only we could find enough ways to kill the microbes that cause them.

Battles still need to be fought, and the report in this issue of the discovery of a potentially new class of antibiotic (see pages 293 and 358) may yield some valuable new weapons. But outright victory looks increasingly unlikely. A surge in strains of resistant bacteria — such as methicillin-resistant *Staphylococcus aureus* (MRSA) — means that an increasing number of bacterial infections cannot be treated effectively. And the growth in the number of emerging



infectious diseases means that the problem is likely to get bigger.

All this is compounded by the withdrawal of many major drug companies from research into antibiotics, and the long timescale for developing new ones (see page 260). Furthermore, persisting in an arms race against microbial resistance may prove futile: it is a race that bacteria are well equipped to win, having evolved mechanisms over millions of years that help them develop resistance to molecules secreted by competing microbes.

Many in the microbiology and immunology communities now believe there is a need for radical new strategies in fighting infectious disease. Developments in immunology and other fields are prompting a convergence towards a more holistic approach that takes into account the complex set of feedback loops between pathogens, host immune systems and our own microbiota.

Better understanding of the host–pathogen interactions at the molecular level may yield answers and open up new ways of thinking about pathogenesis. Rather than always seeking to kill bacteria, for example, molecules that slow their growth or spread may be enough to let the host microbiota and immune system outcompete them, particularly if ways can be found to stimulate or modulate either. Molecules that harnessed our own defences against pathogens would also have the benefit of being less selective for resistance.

Take, for example, *Mycobacterium tuberculosis*, the bacterium that causes tuberculosis. It infects one in ten people in North America

and Western Europe, and one in three worldwide, but only a fraction of these will develop the full-blown disease, with the bacterium dormant or kept in check in the remainder. What tips the balance towards disease? As yet, we have little idea.

Our understanding of the ecology of our own microbiota is limited, as is the molecular basis of host immunity in response to infection. There is still much to understand — until recently, for example, scientists paid little heed to innate immunity, comprising immune cells and secreted molecules that react immediately but rather non-specifically to pathogens. It has only recently been discovered to be much more complex and active than was previously thought.

Such ideas are outlined in an excellent US National Academies report, *Treating Infectious Diseases in a Microbial World* (see <http://fermat.nap.edu/catalog/11471.html>). Infectious diseases have for too long been considered either from the point of view of the microbiologist, with a focus on the pathogen, or from the point of view of the immunologist, with a focus on the host. Basic microbiology, which has lately struggled to win support against competing fields, must not be neglected. But research agencies, universities and scientists should embrace approaches that unite microbiologists and immunologists in the study of infection biology. ■

"Better understanding of host-pathogen interactions at the molecular level may yield answers."

Patents for the people

Peer review comes to the patent office.

Great inventions are driven by good science and technology, but great patent applications succeed by exploiting arcane patent law. This is a paradox of the patent system, and is one reason why it is in danger of being overwhelmed by thousands of complex applications, each trying to stake out the biggest possible piece of the intellectual-property pie.

Both the US and European patent organizations are seriously overstretched: the US Patent and Trademark Office (USPTO) says it has more than a million applications in its backlog; and examiners at the European Patent Office in Munich went on strike last week to protest against their workloads.

Also last week, the USPTO held a meeting in Alexandria, Virginia, to examine a novel approach to the problem. The Community Patent Review project (<http://dotank.nyls.edu/communitypatent>) tries to place the patent process firmly back in the hands of scientists and engineers by asking them for feedback on pending applications.

Under the proposed system, researchers would volunteer to be informed whenever patent applications in their areas of expertise are published. They could then use an electronic bulletin-board to post any prior publications that might be relevant. They could also rank other postings in order of interest. Ideally, the result would be a well-ordered list of publications that the patent examiner could use to determine whether the application is truly innovative.

The idea was the brainchild of Beth Noveck of New York Law School and has won forthright backing from IBM. The computer

corporation, which obtains more patents than any other company, says its support is part of a broader push for patent reform. Now the USPTO is pondering a pilot project of the idea. IBM would volunteer some of its own applications for initial review.

Unsurprisingly, perhaps, some of the patent lawyers who attended the USPTO meeting were unimpressed with the idea. They pointed out grounds for legal challenges to the result of the suggested process: if patent examiners ignored some of the comments they received, for example, they might be accused of "inequitable conduct", which in turn might invalidate the patent. Additionally, comments posted on the site could run foul of current law, which limits communication between patent examiners and outside sources while an application is under review.

Then there is the question of who would bother to participate in the process. For the most part, researchers take part in scientific peer review because the community expects it. No one expects anyone to review patents, and it is conceivable that the only people who come to the site would be rival companies, their lawyers, and cranky 'inventors' hocking cold-fusion reactors and perpetual-motion machines.

But similar pitfalls have not entirely undermined 'open' review systems in other spheres, and the USPTO should at least give the idea a try. Scientific peer review is often credited with keeping research honest, and patent peer review may be able to clean up some of the spurious claims of those looking to profit from vaguely worded applications. And even if it fails, the pilot could assist in the search for fresh approaches for dealing with the backlog. ■

"Patent peer review may be able to clean up some of the spurious claims of those looking to profit from vaguely worded applications."

RESEARCH HIGHLIGHTS

Fly, dragonfly

Proc. R. Soc. Lond. B

doi:10.1098/rsbl.2006.0487 (2006)

Dragonflies, like birds, fatten up and migrate south before winter arrives.

Martin Wikelski, an ecologist at Princeton University in New Jersey, and his colleagues went with them part of the way last autumn.

The team tracked — by aeroplane and car — 14 green darner dragonflies (*Anax junius*, pictured) to which they had glued radio transmitters. They discovered that the bugs' migration is governed by the same few rules as songbird migration: fly during the day, and when the wind is low; take regular days off to replenish your fat supplies; and fly if last night was colder than the night before because this predicts useful northerly winds. One difference, though, is that dragonflies are shorter-lived, so their trips may well be one-way.



C. ZIEGLER

PHYSICAL CHEMISTRY

Water's tangled web

Proc. Natl Acad. Sci. USA

doi:10.1073/pnas.0510593103 (2006)

Do water molecules weave webs or just spin threads? The traditional picture of water has each molecule linked, on average, tetrahedrally to four others in a hydrogen-bonded network. But this model was challenged two years ago by spectroscopic measurements suggesting that the molecules have on average only two hydrogen-bonded neighbours, linking them in chains.

Using X-ray scattering, which probes a wide range of length scales, Teresa Head-Gordon and Margaret Johnson at the University of California, Berkeley, now offer a conciliatory resolution to the controversy. They say the instantaneous local structure

around a water molecule may indeed involve just two hydrogen bonds per molecule — but that this arrangement is constantly in flux, such that the persistent average structure is the standard tetrahedral one.

GEOLOGY

Break away

Geology **34**, 353–356 (2006)

Some geologists think that Russia's Kamchatka peninsula is part of North America. But the geological divide is as great as the cultural one, a new study suggests.

Two great plates of the Earth's crust — the Pacific and North American plates — come together in Russia's far east, but the details of what happens at the junction are murky. New geological studies of Kamchatka, including uplifted layers of ancient seashores (pictured

below), suggest that most of the peninsula rests on its own, separate crustal block.

Seismic data from two earthquakes that occurred in the region in late April could help to further clarify the picture, says the team that carried out the study, comprising researchers in China, the United States and Russia.

NANOTECHNOLOGY

Sort it out!

Science **312**, 910–914 (2006)

Tiny tubular proteins can now be sorted into different containers as they 'crowd surf' over the backs of kinesin motor proteins.

Cees Dekker and colleagues at Delft University of Technology, the Netherlands, coated sub-micrometre-sized channels running through a slice of silica with kinesins. These proteins, which normally carry cargo along microtubule filaments within cells, have been used to transport microtubules on chips before. But enclosing them in channels has offered new control: using a voltage, Dekker's group could direct the surfing microtubules down either the left or right fork of a Y-shaped junction into a collecting pool.

MICROBIOLOGY

Surprise attack

Proc. Natl Acad. Sci. USA **103**, 6724–6729 (2006)

The pneumonia-causing bug *Mycoplasma pneumoniae* uses an unsuspected weapon to attack the lungs.

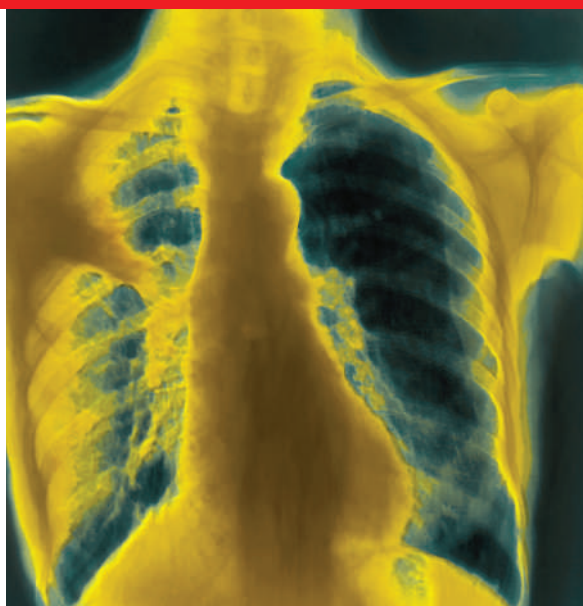
T. R. Kannan and Joel Baseman of the



J. BOURGEOIS

University of Texas Health Science Center in San Antonio show that *M. pneumoniae* produces a harmful cytotoxin. Previously it had been assumed that the lung damage typical of pneumonia (pictured right in X-ray) was caused by the body's immune response to the infection because there were no known mycoplasma toxins.

Kannan and Baseman identified the toxin, an ADP-ribosylating enzyme, using a human lung protein as bait to bind it. They showed that the toxin altered certain mammalian proteins, leading to vacuolization, or the formation of harmful membrane structures, inside cells. They also identified antibodies to the toxin in blood from pneumonia patients.



Montreal in Canada, John Westwick of Odyssey Thera in San Ramon, California, and their team.

The researchers used an assay in which two proteins, each fused to a fragment of a fluorescent protein, trigger a yellow signal when they come together. They created 49 such assays for proteins involved in diverse cellular processes and tested whether 107 known drugs interfere with these protein interactions.

The screen revealed that some drugs have a previously unrecognized ability to block cell proliferation — and identified a panel of tests that could predict whether a new drug can stop cell division and potentially fight cancer.

CNR/SPL

ECOLOGY

Present and correct?

Ecography 29, 129–151 (2006)

Ecologists need to model species' distributions to predict, among other things, the impact of climate change on biodiversity. Jane Elith of the University of Melbourne, Australia, and Catherine Graham of Stony Brook University, New York, and their colleagues provide tips for making the best use of limited data.

They evaluated the predictive ability of 16 different methods for modelling species' distributions, given data that say only where a species has been found, rather than where it has not. Such presence-only data can be gleaned from museum records.

They tested the models' predictions for 226 species against independent data sets containing both presence and absence information. They conclude that newer methods generally outperform established tools such as GARP and BIOCLIM.

NANOTECHNOLOGY

New ice age

Phys. Rev. Lett. 96, 166103 (2006)

Physicists investigating the effect of water vapour on friction at the nanoscale say that ice may form at room temperature.

The capillary forces caused by the condensation of water vapour are a major issue in many nanoelectromechanical systems. This motivated K. Jinesh and Joost Frenken from Leiden University in the Netherlands to explore the movement of a sharp tungsten tip across a graphite surface in a humid atmosphere. They noticed that the water was acting like glue rather than a lubricant.

The results can be explained in terms of the tip writing a temporary line of ice on the surface, but the substance differs from normal ice in ways that are not yet understood.

DRUG DISCOVERY

Seeing all sides

Nature Chem. Biol. doi:10.1038/nchembio790 (2006)

A new way to reveal unwanted or unexpected drug side effects has been demonstrated by Stephen Michnick of the University of

CELL BIOLOGY

Crosstalk

Nature Cell Biol. doi:10.1038/ncb1418 (2006)

The first large-scale, systematic analysis of interactions between the signalling pathways that mediate a cell's response to external signals has revealed extensive crosstalk. But the connections are not, as some people feared, so numerous that they will be impossible to unpick.

Rama Ranganathan and his colleagues at the University of Texas, who are part of the Alliance for Cellular Signalling, studied the effect of 22 extracellular signalling molecules, such as hormones and cytokines, on a macrophage cell line.

They measured crosstalk by monitoring, for all possible pairwise combinations of these ligands, various parameters in the signalling pathways downstream of the ligands' receptors. These parameters included calcium levels and the phosphorylation of signalling proteins.

JOURNAL CLUB

Albert Bosma

Marseille-Provence Astronomical Observatory, France

An astronomer explains how galaxies keep him drawn in to the dark-matter debate.

Ours is a strange Universe. It seems to contain far more dark stuff than ordinary matter.

In the 1970s, when I was a PhD student, galactic dynamics played an important role in establishing the existence of dark matter. For

my thesis, I studied the rotation of spiral galaxies. The rotation profiles implied dark matter — some unseen thing exerting a gravitational pull kept the rotation speed high even beyond the galaxies' visible edges.

Today, rotation curves pose a strong challenge to the most popular model of how dark matter fits into the Universe, Λ cold dark matter (Λ CDM) theory. For galaxies, numerical simulations based on Λ CDM theory predict that their dark-matter density peaks towards the centre. But this doesn't

square with rotation data that suggest the dark matter in so-called 'low surface brightness' galaxies is distributed evenly through the core.

Some say the rotation data are problematic. Others suggest ways to reconcile the observations with theory — perhaps a bar of stars in the galactic disk somehow erases the central dark-matter peak.

We might see the kinematic effects of bar-like features with better rotation data and deep images of the faint disk. But so far, our newer results seem to uphold the discrepancy.

Recent galactic simulations with an impressive billion particles of dark matter (J. Diemand *et al.* *Mon. Not. R. Astron. Soc.* 364, 665–673; 2005) actually bolster the peak theory.

Yet, simulations are not theory. We don't really know what sets the dark-matter profile — it may be some interaction with real matter, with the gas and stars in a growing galaxy. We need far more elaborate modelling than is possible now. By the time it is, I hope to have even better data to keep up the challenge.

NEWS

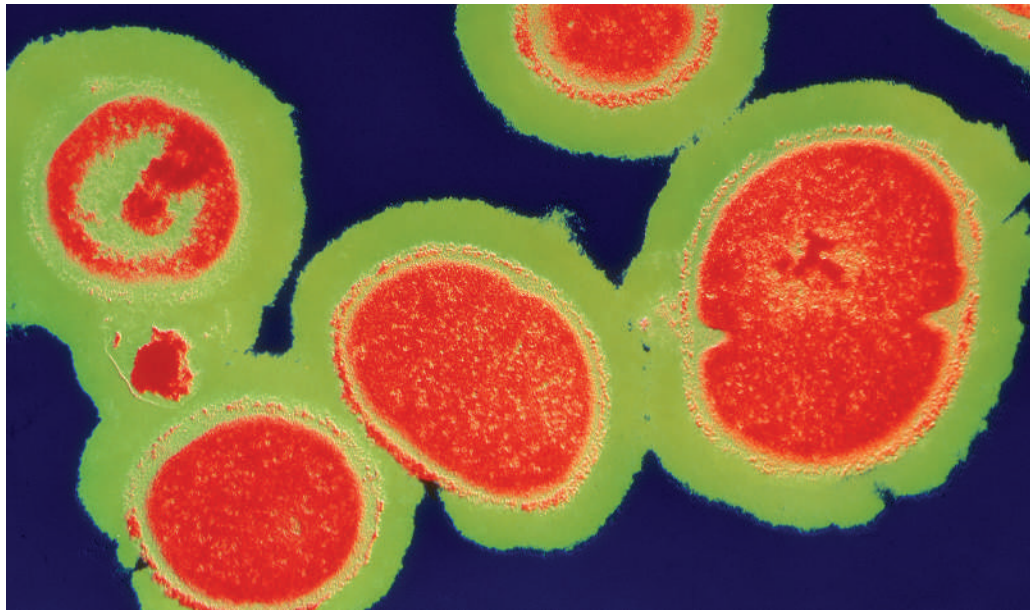
Antibiotic faces uncertain future

A promising new antibiotic is generating both excitement and despondency. It is the first new chemical class of antibiotic to be found in more than two decades, but experts fear that the hurdles to turning the compound into an effective commercial drug could mean that it ends up collecting dust on a shelf.

The drug, called platensimycin and reported on page 358, kills several of the major drug-resistant bacteria that plague hospitals. Among them are methicillin-resistant *Staphylococcus aureus* (MRSA) and bacteria resistant to vancomycin, one of the last lines of antibiotic defence.

Only two other new chemical classes of antibiotic have been discovered and approved for use since the early 1960s: daptomycin and linezolid. Most antibiotics used today were discovered in the 1940s and 1950s, and newer versions have mainly been made by chemically nipping and tucking these compounds. They generally kill bacteria by blocking the production of proteins, DNA or the bacterial cell wall.

Platensimycin has a novel chemical structure and works differently from other commercially available antibiotics by crippling FabF, a bacterial enzyme involved in manufacturing fatty acids. It thus stops bacteria from making the fatty cell membranes they need to



An antibiotic has been discovered that kills the deadly methicillin-resistant *Staphylococcus aureus*.

grow. Two commercially available antibiotics, triclosan and isoniazid, target another enzyme involved in fatty-acid synthesis, but these do not kill the same array of bacteria.

The researchers at Merck Research Labora-

tories in Rahway, New Jersey, who discovered platensimycin, did so by reviving and tuning an established strategy. They searched for naturally existing compounds that microbes typically make to kill neighbouring bacteria. This

K. LOUNATMAA/SPL

Neanderthal DNA yields to genome foray

NEW YORK

The first nuclear DNA sequences from a Neanderthal (*Homo neanderthalensis*) have been reported. The results should provide clues about when certain diseases, or traits such as hair or skin colour, arose. They also have geneticists excited about the idea of sequencing a Neanderthal genome.

Svante Pääbo, a palaeogeneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, began his Neanderthal Genome Project about two years ago. He and his team have probed 60 Neanderthal specimens from museums for hints that the DNA might have survived millennia of degradation. The species lived across Europe and western Asia from

300,000 to around 30,000 years ago, with the first specimen found in 1856 near Dusseldorf, Germany.

Two of the specimens showed promise, and on 12 May Pääbo's team reported at the Biology of Genomes meeting at New York's Cold Spring Harbor Laboratory that they had managed to sequence around a million base pairs of nuclear DNA — around 0.03% of the genome — from one of them. This is a 45,000-year-old male specimen found in Vindija Cave outside Zagreb, Croatia.

Typically, DNA to be sequenced must be cloned in bacteria to produce large enough amounts for study. But because the Neanderthal DNA had broken down into tiny fragments, Pääbo and his

colleagues used a new sequencing technique, developed by 454 Life Sciences in Branford, Connecticut, that allows genetic fragments in an emulsion to be sequenced directly in tiny wells. They are now analysing the results to work out how the different fragments fit together so that they can be compared with the modern human genome sequence.

One finding so far is that the Neanderthal Y chromosome is substantially more different from human and chimp Y chromosomes than are other chromosomes. This suggests that little interbreeding occurred, at least among the more recent Neanderthal species.

Edward Rubin, director of the Joint Genome Institute in Walnut Creek, California, works with Pääbo.

The two are also working to sequence Neanderthal DNA by the traditional method. James Noonan, a postdoc in Rubin's lab, reported at the Cold Spring Harbor meeting that preliminary analysis of the 75,000 base pairs sequenced so far shows that Neanderthals diverged from the lineage that led to modern humans about 315,000 years ago — around the time that had been thought. *Homo sapiens* is known to have evolved at least 200,000 years ago (I. McDougall, F. H. Brown and J. G. Fleagle *Nature* 433, 733–736; 2005).

Back in 1997, Pääbo reported sequencing the first mitochondrial DNA from a Neanderthal (M. Krings *et al. Cell* 90, 19–30; 1997). That sequence also suggested that Neanderthals split from the



RECORD-BREAKING LASER IS HOT STUFF

The record for the fastest rise in temperature has just been topped.

www.nature.com/news

S. JUODKAZIS ET AL.

technique originally yielded many potent antibiotics, but was abandoned by many drug companies as the rate of return diminished and has been replaced by chemical methods to generate novel synthetic molecules.

Traditionally, researchers have tested whether a tiny amount of chemical can kill off a circle of bacteria growing on a plate. But by first making the bacteria more vulnerable, the Merck researchers searched for natural products that might have been missed in such conventional assays. They engineered *S. aureus* bacteria to make less of the FabF enzyme than normal, using a snippet of RNA to block the production of this protein (K. Young *et al. Antimicrob. Agents Chemother.* 50, 519–526; 2006).

The researchers then tested whether any of 250,000 natural-product extracts could block the growth of the disabled strain — and pulled out platensimycin, a small molecule made by a strain of *Streptomyces platensis* bacteria in a South African soil sample. “It allowed us to find a needle in a haystack, and something we might have missed with another type of screen,” says Stephen Soisson, one of the lead researchers.

Other scientists are delighted with the new compound, and say it could provoke renewed interest in natural compounds and in attacking the fatty-acid synthesis pathway. “From a scientific point of view it is what you want — a new

class against a new target,” says Steven Projan of Wyeth Research in Pearl River, New York.

But “the next steps are fraught with danger,” warns microbiologist Carl Nathan of Weill Medical College of Cornell University in New York. “The obstacles are truly formidable.”

Platensimycin could stumble at one of many scientific, regulatory or financial hurdles. One concern is that it may be unstable in the body, because the research team had to infuse it continuously into mice to rid them of a *S. aureus* infection. The chemical might need extensive modification to make it more stable, and could prove useless if it has toxic side effects.

Testing antibiotics in clinical trials is also tough and expensive. This is partly because the US Food and Drug Administration (FDA) does not have clear guidelines for approving new antibiotics, says Frank Tally, chief scientific officer of Cubist Pharmaceuticals in Lexington, Massachusetts, which developed daptomycin. Even if platensimycin is approved for human use, it will probably meet the fate of its predecessors, as bacteria rapidly acquire resistance to it. “It might make it to the clinic and then fade because of drug resistance,” Nathan says.

With such a daunting task ahead, some microbiologists fear there is little incentive for Merck to develop the drug. Not only is the process hugely expensive, but the potential market for a new antibiotic — which is often

used sparingly and administered for only a week or two — is small. This has prompted many pharmaceutical giants to cut back their antibiotic research programmes in recent years.

The Merck researchers refuse to spell out their plans for platensimycin, although they hint that they are working on it. “We are still in the business of doing antibiotic research, and we wouldn’t be if we didn’t want to develop our own compounds,” says Sheo Singh, another of Merck’s lead researchers.

Several solutions have been proposed to keep new drugs flowing into the clinic as microorganisms overcome old ones. The Infectious Diseases Society of America proposed an array of regulatory measures in 2004 that might entice drug companies to develop antibiotics, such as revised FDA standards for the approval of antibiotics and tax breaks for work on such drugs (see *Nature* 431, 892–893; 2004). Some researchers are turning to alternative measures to fight life-threatening bacteria, such as microbe-killing bacteriophage and vaccines.

But many workers still think that novel antibiotics remain the best tried and tested way to combat microbes. They say platensimycin’s discovery highlights the fact that compounds are still waiting to be found. “With a paper like this, people might start to reinvest in this area,” Tally says, as it shows that new molecules can be found in well-known sources.

Helen Pearson

“Platensimycin’s discovery highlights the fact that compounds are still waiting to be found.”

common lineage well before modern humans evolved and may not have contributed any mitochondrial DNA, at least to modern humans.

But the more extensive nuclear DNA sequences should pin down the timing of the split more precisely, and comparing genes for particular traits could help researchers work out which characteristics were shared by Neanderthals, and when such traits arose. Such comparisons could also confirm whether Neanderthals did contribute isolated genes to the human lineage. For example, John Hardy, a geneticist at the National Human Genome Research Institute in Bethesda, Maryland, has hypothesized that Neanderthals may have contributed a gene that is linked to several neurodegenerative diseases, because it is found in people of European ancestry, where

the Neanderthals lived. Proving that theory would require finding this version of the gene in the Neanderthal genome.

Researchers are confident about the prospects for sequencing much more Neanderthal DNA. “Our goal is to sequence a large amount of the Neanderthal genome,” says Pääbo. That task will probably require identifying new fossils. “We should do it,” agreed Francis Collins, director of the National Human Genome Research Institute, after attending Pääbo’s lecture on 12 May.

Rubin says he envisages creating a bank of Neanderthal DNA, so that representative samples can be compared with the thousands of *H. sapiens* genomes that are expected to be sequenced in the future. “In ten years, we hope to have ten Neanderthal genomes,” says Rubin. ■
Rex Dalton



Neanderthal (right) DNA should shed light on the human genome.

G. J. SAWYER & B. MALEY; PHOTO D. FINNIN/AMNH

How does a painkiller harm the heart?

It was Garret FitzGerald, the chair of pharmacology at the University of Pennsylvania, who first proposed that the painkiller Vioxx and its cousins might cause heart attacks by tipping a crucial hormonal balance. So neither of two attention-grabbing data sets published this month has come as a great surprise to him, despite the implications the data are likely to have for the 11,500 lawsuits from former patients that dog the drug's manufacturer, Merck.

One study, published in the *Canadian Medical Association Journal*¹, shows that Vioxx, now withdrawn, caused heart attacks within days of users first taking it. The other data, released last week by Merck, suggest that subjects remained at increased risk of heart attack and stroke for a year after they stopped taking the drug².

"These data are extremely interesting if you're a lawyer, but scientifically and mechanistically they show exactly what you'd expect," says FitzGerald.

Others disagree. Nobody expected the Merck results, says Steven Nissen, a cardiologist at the Cleveland Clinic in Ohio. "We all thought that if you stopped the drug, the risk would go away. These data suggest there may be more longstanding injury to blood vessels."

The difference of opinion reflects uncertainty over exactly how heart attacks and strokes are caused by Vioxx (rofecoxib) and related drugs, which all work by inhibiting an enzyme known as cyclooxygenase-2. It is widely accepted that these 'COX2 inhibitors' exert their effects in part by suppressing prostacyclin — a product of the COX2 enzyme. Prostacyclin plays a key role in mediating pain and inflammation: the drugs were developed to combat inflammatory diseases such as arthritis.

Off balance

But studies in mice have shown that prostacyclin also acts as a general restraint on platelet activation, and on factors that accelerate hardening of the arteries (atherosclerosis), boost blood pressure and activate blood clotting. Key among these factors is a molecule called thromboxane. Crudely put, prostacyclin and thromboxane normally balance each other's opposing effects. Introducing a COX2 inhibitor can tip the balance dangerously toward thromboxane, raising blood pressure,



Pulled: after the scandal of withdrawal, Vioxx is unlikely to go back on sale. But its sister drugs may have a future.

possibly hardening arteries and certainly promoting heart-attack and stroke-causing clots in some people.

The Canadian study, which found that Vioxx's adverse effects kicked in almost immediately, fits neatly with this theory. There is an established link between prostacyclin suppression and excessive clotting, points out Eric Topol, a geneticist at Case Western Reserve University in Cleveland. "If you start to suppress prostacyclin, you could see a clot form relatively quickly. Why would you need months or a year to see that?"

The Merck data are more difficult to explain. FitzGerald says they fit with his work in mice, which shows that COX2 inhibitors can cause atherosclerosis — not an injury that would recede rapidly once the offending drug is stopped. But this has not been demonstrated in humans.

"I don't think it's well-established" that Vioxx causes atherosclerosis in people, says Nissen. "But Garret's been right about a lot. He may be right about this."

Frank Ruschitzka, a cardiologist at the University Hospital in Zurich, who has studied Pfizer's Celebrex, another COX2 inhibitor, says it is a mistake to draw any mechanistic conclusions from Merck's report. The study it is drawn from, he notes, was not designed to

examine heart attacks³. "We don't know the mechanisms," he says. "So everyone involved should just lower their voices."

Ruschitzka has designed a trial of 20,000 subjects to be launched later this year, with Nissen as principal investigator. It will compare the incidence of heart attacks, strokes and cardiovascular deaths in subjects taking Celebrex (celecoxib) with those using the traditional anti-inflammatories ibuprofen and naproxen. Only such a huge, randomized, head-to-head controlled trial — paid for by Pfizer at a cost of hundreds of millions of US dollars — will allow the science of the side effects to be truly sorted out, he says.

Ruschitzka believes that more mechanisms are at work than the balance-tipping theory admits. For instance, in one recent paper, he showed that Celebrex reduces expression of a protein that initiates coagulation⁴. Vioxx doesn't have this effect. "If FitzGerald tells us it's just

one mechanism, he's wrong," says Ruschitzka.

Others, including Topol, believe that a minority of individuals may have a genetically determined sensitivity to Vioxx. But this could be pinned down only by a genome-wide study comparing those who develop clotting with those who do not. And that's not a study that Pfizer or Merck are rushing to sponsor.

"We don't know the mechanisms, so everyone should just lower their voices."

The quest to clear these murky waters isn't just academic, even if Vioxx itself is unlikely ever to be resurrected. It is hoped that other COX2 inhibitors may eventually treat diseases as diverse as colon

cancer and Alzheimer's. And it is therefore vital to flesh out the mechanism of their cardiac dangers, so the risk-benefit equation is understood before the drugs saturate new markets.

As for former Vioxx patients, Nissen says they should all be closely monitored for at least a year. He tells his own patients that their risk of heart attack has been almost doubled, and that "they should do everything they can to control their risk factors".

Meredith Wadman

1. Lévesque, L. E., Brophy, J. M. & Zhang, B. *Can. Med. Assoc. J.* **174**, doi:10.1503/cmaj.051679 (2006).
2. Merck APPROVe report (2006); www.merck.com/newsroom/pdf/APPROVe_Study_Summary_Report.pdf
3. Bresalier, R. S. *et al. N. Engl. J. Med.* **352**, 1092-1102 (2005).
4. Steffel, J. *et al. Circulation* **111**, 1685-1689 (2005).



Suspected: nearby farmed geese, not migrants from afar, might have infected birds at Qinghai Lake.

CHINA PHOTOS/GETTY

Blogger reveals China's migratory goose farms near site of flu outbreak

The hypothesis that migratory birds are responsible for spreading avian flu over long distances has taken another knock. Last year, an outbreak of the deadly H5N1 strain in thousands of migratory birds at Qinghai Lake in western China provided what seemed the first firm evidence for the idea. Because the lake is so remote, experts assumed infected birds had flown up from southern China.

But it has now emerged that, since 2003, one of the key migratory species affected, the bar-headed goose, has been artificially reared near the lake. The breeding farms — part of an experimental programme to both domesticate the birds and release them to repopulate wild stocks — raise the possibility that farmed birds were the source of the outbreak.

Roy Wadia, a World Health Organization (WHO) spokesman in Beijing, agrees that, if confirmed, the finding is "important", as changing the breeding practice might help control the infection.

Yi Guan, a virologist at the University of Hong Kong, co-authored a *Nature* paper last July that suggested migrating birds caused the outbreak (see *Nature* 436, 191–192; 2005). Guan says he had heard rumours of the programme when he submitted his paper, but couldn't confirm them.

There is no proof that China's breeding programme caused the Qinghai outbreak, but it does raise questions, he says. "The cultivation of bar-headed geese increases the chance for these birds to mix with infected domestic poultry."

Ironically, the breeding programme was revealed by Chinese press agencies reporting on the government's efforts to boost agriculture and the environment in the region ahead of the opening of the Qinghai-Tibet railway in July; the railway is expected to promote tourism and economic growth.

Richard Thomas of BirdLife

International in Cambridge, UK, spotted the press cuttings, and posted English translations to a blog (www.drmartinwilliams.com). Whether farmed migrant birds caused the outbreak or not, it's a "cautionary tale", says Ken Shortridge, a veteran avian-flu researcher in China. He argues that such a programme does not sufficiently take into account the threat of H5N1.

"Whether farmed migrant birds caused the outbreak or not, it's a cautionary tale."

The idea that migrating birds didn't carry the virus to

Qinghai after all would fit with other recent evidence. Juan Lubroth, a senior animal-health officer at the UN Food and Agricultural Organization (FAO), says he is now sceptical that migrants can carry the virus over long distances. For example, the current spring migration from Africa to Europe is almost over, with no sign of outbreaks. The FAO has also checked 20,000 wild birds in Africa and found no H5N1. ■ Declan Butler

ON THE RECORD

“There’s no way that I’m ever going to be able to serve liquid nitrogen margaritas.”

Bartender Eben Klemm recognizes the limits of ‘molecular mixology’, the use of science to create innovative cocktail experiences.

“I feel that I have been in a war and under siege for six years.”

Christopher Hall reacts to the jailing of four UK animal-rights campaigners who terrorized his family’s farm, which until January bred guinea pigs for research.

Source: *New York Times*, BBC

SCORECARD

**Flu fighters**

A study of two old, generic drugs contradicts previous reports that they don’t work against the deadly H5N1 bird-flu virus, boosting hopes that amantadine and rimantadine might help control a pandemic.

**Penguins**

Researchers are hunting for the source of a mysterious oil spill that has killed 100 Magellan penguins off the southern coast of Argentina.

NUMBER CRUNCH

Cynics often suggest that clinical trials funded by drug companies have a knack of supporting new treatments over old. But there may be more to this than gut feeling, reveals a survey looking at trials that tested new cardiovascular therapies against established treatments.

49% of 104 trials sponsored by not-for-profit entities came to the conclusion that the newer treatments were better.

56.5% of 62 trials sponsored by partnerships between for-profit and non-profit entities favoured new treatments.

67.2% of 137 trials sponsored by for-profit funders found evidence supporting the use of the new treatment.

Source: Ridker, P. M. & Torres, J. *J. Am. Med. Assoc.* **295**, 2270–2274 (2006).

PHOTOFUSION PICTURE LIBRARY/ALAMY



Panel beaters: could low-carbon energy be surprisingly affordable?

Economists claim carbon cuts won't break the world's bank

Transforming the world’s energy industry to stop the flood of greenhouse gases into the atmosphere might actually be quite cheap.

Figures of tens of trillions of dollars are often cited, and used to question whether measures such as the Kyoto Protocol, which attempts to limit carbon emissions, are too expensive. But according to a suite of economic models released late last month, the costs of stabilizing carbon dioxide levels could be tiny — equivalent to setting back the growth of global GDP (gross domestic product) by less than 1% over 100 years; global GDP generally grows 2–3% each year. In some cases, the right policies for limiting carbon emissions could even create a surprising win-win situation, leading to the stabilization of greenhouse gases and an increase in global wealth.

It is a controversial conclusion with which not everyone agrees, but which the modellers say should be food for thought for policy-makers. “If we prepare properly and acknowledge that carbon will be constrained, it will be relatively cheap,” says Michael Grubb, a climate-policy expert at Imperial College

London. “But only if we do the right things.”

The models simulate a complex issue in economics: how government climate policies such as research investment or greenhouse-gas regulation can bring about technological development. It is obvious that technologies evolve, but the processes involved have been factored into economic models only since the late 1990s, in part because it is difficult to untangle how advances occur. The Innovation Modelling Comparison Project, published in a special issue of *The Energy Journal*, is a two-year

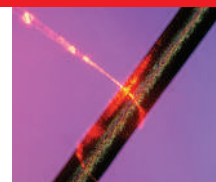
effort involving eleven different models that represent the latest thinking on the problem.

The results are striking. Nine of the models predict that stabilizing carbon dioxide levels at

450 parts per million, widely seen as the most ambitious target worth discussing, would set back global GDP by less than 0.5% or so by 2100 (the other two produced figures of 2.1% and 6.2%). In each scenario, the regulation of greenhouse-gas emissions persuades the private sector to shift investment into low-carbon technologies, which then become competitive with traditional energy sources.

In some cases, this shift in investment

“Reducing greenhouse gases will be relatively cheap — but only if we do the right things.”

**PHYSICS NEWS**

Get breaking news on the physical sciences from our website.

www.nature.com/news/channels/physicsciences.html

L. TONG/HARVARD UNIV.

stimulates growth and actually boosts overall wealth. At least, that's the conclusion of two of the models — one developed at the University of Cambridge, UK, and the other at the Fondazione Eni Enrico Mattei, a centre for sustainable-development research in Italy. These models suggest that stabilization policies would give an added boost to global GDP of up to 1.7% over 100 years. They assume such climate policies will bring about side benefits, such as increased investment in new technologies.

Ottmar Edenhofer, an economist at the Potsdam Institute for Climate Impact Research in Germany who edited the issue along with Grubb and others, says the new estimates of lost global GDP are significantly lower than previous ones, which put the range at 3–15%. They suggest the price will be a lot lower, agrees Terry Barker, an economist who helped develop the Cambridge model, especially as costs will be spread over 100 years.

The models are likely to influence the next report from the Intergovernmental Panel on Climate Change, due for publication next year. The authors hope the results will then filter through to governments. They say the cheapest stabilization route can only be achieved if industries are given a strong signal that carbon emissions will continue to be restricted — and that means the United States must join a future version of the Kyoto Protocol. Europe also needs to do more, say the authors, particularly in terms of investment in energy technologies, where it lags behind the United States.

But some economists are wary of the results. Jae Edmonds of the Pacific Northwest National Laboratory in Richland, Washington, describes the models as a valuable “intellectual experiment”. But he questions the fact that most of the models emphasize learning-by-doing — a process by which technologies become cheaper as industry learns how to produce them more efficiently.

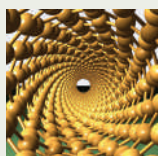
Edmonds points out, for example, that many advances in hybrid vehicles have been made possible by developments in areas outside the automotive industry. It is hard to distinguish these advances from those due to learning-by-doing or research by automotive firms. If advances are attributed exclusively to one mechanism, Edmonds warns that the benefits of the process could be exaggerated and costs underestimated: “We don't necessarily have a good handle on how important different factors are and how they interact.”

Jim Giles

TOP FIVE IN PHYSICS

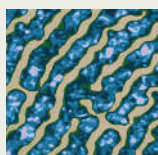
Are you working on the hottest topic in your field? Many scientists may think so, but it has been a tough assertion to prove — until now, that is. A German physicist has devised a way of answering the ‘Hot or not?’ question for his discipline. If it stands up to scrutiny, it could be used to rate topics across the sciences. In physics, the results show that hotness — measured by a parameter known as m — correlates well with the promise of future wealth... and that promise is greatest in nanotechnology.

12.85 Carbon nanotubes



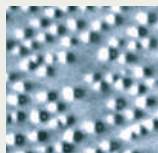
Super-strong materials and blisteringly fast electronic circuits: the potential applications of these tiny carbon tubes, discovered in 1991, are so enticing that everyone is pouring money into the field.

8.75 Nanowires



Less well studied than nanotubes, but the possible uses are similar. Nanowires could eventually prove more useful than nanotubes, because their chemistry is easier to tailor and they can be used to create nano-sized lasers.

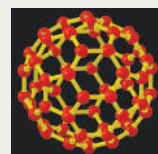
7.84 Quantum dots



Another nanotechnology with a huge range of potential applications. These tiny specks of semiconductor material, measuring as little as a few nanometres across, have already been used to create dyes for cell biologists and new kinds of

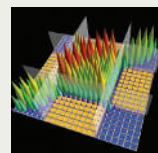
laser. Physicists hope they might one day form the basis of a quantum computer.

7.78 Fullerenes



These spheres of carbon atoms are attracting significant research interest. But the latest ranking rewards newness, so the topic may have slipped down the list because it predates nanotubes by around six years. The discovery of fullerenes earned a Nobel prize and spawned studies of numerous potential uses, such as drug delivery agents.

6.82 Giant magnetoresistance



Not a new topic, but still hot because of its economic importance. Modern hard disk drives were made possible by the discovery of giant magnetoresistant materials, which show marked falls in electrical resistance — more than around 5% — when a magnetic field is applied. Researchers are now aiming to make hard disks even more powerful.

How the topics were ranked

The ranking is an extension of a recently proposed system for rating the research output of individual scientists. The h index uses the highest number of papers a scientist has published that have each received at least that number of citations: for example, a researcher with an h of 50 has written 50 papers that have each had at least 50 citations (see *Nature* **436**, 900; 2005). The index has attracted interest from bibliometricians, but was first described only last year and has yet to be studied in depth.

Michael Banks, a PhD student at the Max

Planck Institute for Solid State Research in Stuttgart, Germany, totted up the h index for physics topics, rather than people, and then calculated the parameter m by dividing by the number of years over which the papers involved had been published (www.arxiv.org/abs/physics/0604216).

Jim Giles

Do you agree with the winners? Tell us what you think on our news blog

► blogs.nature.com/news/blog/2006/05/top_five_in_physics.html

FROM START: S. HEINZE/H. JAEGER & W. LOPES; NIST; NO CREDIT; ORNL/B. BUTLER

Who is Mark Myers?

That's what many US geologists are asking in the wake of an announcement that President George W. Bush will nominate Myers to head the US Geological Survey (USGS). The agency spends nearly \$1 billion each year on research, including earthquake monitoring, mapmaking and even biological studies.

Myers has a PhD in geology and has spent much of his career in Alaska, working for oil companies and for the state — sometimes alone in remote locations, armed with a shotgun in case of grizzly bears. "Never had to use it," he says. "I keep a clean camp." He served in the Air Force Reserve for 26 years, for a time in intelligence in Alaska's far north.

If confirmed by the Senate, Myers would be the first USGS director in decades to come neither from academia nor from within the agency. His history has made many academic geologists nervous, says Charles Groat, the agency's previous director. "The biggest question is, since he is so identified with oil and gas, what is his

agenda?" says Groat, now at the University of Texas, Austin.

Myers worked most recently as head of Alaska's Division of Oil and Gas. In the past he has supported drilling for oil and gas in the Arctic National Wildlife Refuge — a protected region of Alaska. And this has spooked some environmentalists. But if he gets the USGS job, Myers says, he would stay out of any decision making: "My job is strictly to provide the data, to help people understand the data and its limitations."

Last autumn, Myers resigned in protest after Alaska's governor dismissed Tom Irwin, then natural resources commissioner for the state. The dispute was over a natural-gas pipeline deal between Alaska and major oil companies. Linda Rowan, government affairs

director at the American Geological Institute in Alexandria, Virginia, says she is reassured by the incident: "If he's willing to resign from his very prominent position in Alaska, he's not really going to be political."

That is, of course, the big question. The Bush administration has a reputation for manipulating science to fit its ideology and goals. But Myers says that during his job interview with the White House personnel office, "one of the key things I was asked is: how do you maintain the objectivity of the science in a political environment? I was very encouraged by this question." He says the only politically directed question put to him was: "Would you support the president?" — to which he said yes.

Emma Marris



Mark Myers may soon head the US Geological Survey.

Australian budget builds on health research

Medical and health research is a major beneficiary of this year's Australian government budget, but basic science has been short-changed, critics say.

The budget, announced on 9 May, injects an additional A\$500 million (US\$380 million) over four years into the National Health and Medical Research Council, the nation's main funding agency for biomedical research, on top of its annual A\$490-million budget. Medical-research infrastructure will be boosted by A\$213 million, and a new research fellowship scheme of A\$170 million over nine years will support up to 60 senior researchers.

Although universities will get an additional A\$560 million over five years, this is primarily to fund new student places and infrastructure developments. "There is new funding for physical infrastructure, but I'm disappointed the budget doesn't reflect an investment in intellectual infrastructure," says geophysicist Kurt Lambeck, president of the Australian Academy of Science.

Trophy hunter bags crossbred bear

An unusual-looking bear shot by a hunter in northern Canada has been confirmed as a grizzly-polar bear hybrid, the first ever found in the wild. Some conservationists have speculated that climate warming might drive grizzlies north where they could mate with polar bears.

DNA testing showed that the animal — a male shot by a licensed sport hunter near Banks Island on 16 April — was the

offspring of a male grizzly and a female polar bear. The bear, which has been dubbed both a 'pizzly' and a 'grolar', is mostly white but has brown eye patches and the large paws and claws of a grizzly. Local officials have given the hide back to the hunter — after earlier speculation that he could have been liable to prosecution for shooting an animal not covered by his licence.

NASA cargo to hitch a ride on Indian spacecraft

Two payloads from NASA will fly on board Chandrayaan-1, the first unmanned Indian spacecraft to the Moon, which is scheduled for launch in early 2008.

A synthetic aperture radar developed by Johns Hopkins University in Maryland, will look for water in the Moon's polar regions, and an instrument designed by the Jet Propulsion Laboratory in Pasadena, California, will map and characterize minerals on the lunar surface.

The Indian Space Research Organization will fly the equipment for free, and NASA will share its data with Indian scientists, under an agreement signed on 9 May.

Four other instruments — from the United Kingdom, Germany, Sweden and Bulgaria — were also selected out of 16 proposals from abroad. They will join India's own payloads, which include a terrain-mapping camera and a Moon-impact probe that would test future landing technologies.

Chandrayaan-1 will scrutinize the lunar surface for two years from an altitude of 100 kilometres. NASA's participation is seen as a sign of revival of Indo-US cooperation in space that was interrupted by sanctions imposed by Washington following India's nuclear tests in 1998.

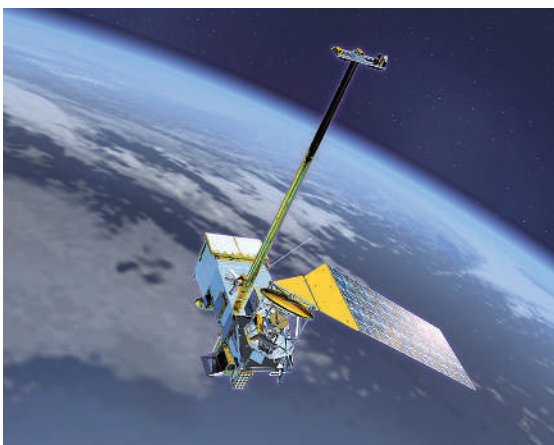
Prize aims to fuel clean-energy technology

Hydrogen-powered vehicles (pictured) are picking up momentum in the US Congress. The House of Representatives passed a bill on 10 May that would set up a prize to reward innovators who create hydrogen technology to provide alternative energy sources. The Senate must now approve the idea before it becomes law.

Modelled after last year's privately funded \$10-million Ansari X Prize, which led to the first private suborbital spacecraft, the H Prize aims to encourage breakthrough technologies. The tiered award would run from 2007 to 2016. Every two years, it would grant four prizes of up to \$1 million for innovations in hydrogen storage, distribution, production and use; on alternate years, a prize of up to \$4 million would go to a working prototype in these areas. There is also a grand prize of \$10 million, although winning criteria have yet to be decided. Backers hope the prize will be funded by a combination of the federal government and the private sector.



S. YEATER/AP



NOAA

Poor forecast: the US project to build six NPOESS weather satellites is running late and over budget.

America's merged satellite system hits stormy weather

It seemed like a good idea — merge the US civilian and military weather-satellite systems, avoid duplication, and save money. But 12 years into the National Polar-orbiting Operational Environmental Satellite System (NPOESS), the dream has turned into a nightmare.

Originally budgeted at \$6.5 billion for six satellites, NPOESS is now expected to cost twice as much, and its planned operational readiness date of 2009 has slipped three years. A government audit released last week detailed a slew of management blunders, including awarding too-generous contracts to industry and reacting too slowly to signs of trouble.

Alarmed lawmakers say that the United States could face a gap in weather forecasting ability. They add that problems with the NPOESS programme could bankrupt the National Oceanic and Atmospheric Administration (NOAA), which manages the system for a partnership that includes the defence department. Some congressional Democrats are calling for NOAA chief Conrad Lautenbacher to step down because of the problem.

Glaxo shareholders win protection from extremists

British drug giant GlaxoSmithKline (GSK) obtained an injunction from the High Court in London last week, forbidding animal-rights activists to contact or harass individual GSK shareholders, or publish their names and addresses.

On 8 May, hundreds of shareholders began receiving anonymous letters demanding that they sell their stock in the company or face having their names and addresses published on the Internet. More than 400 concerned shareholders contacted GSK.

The activists say in their letter that they want GSK to sever links with the animal-research firm Huntingdon Life Sciences. GSK's predecessor, GlaxoWellcome, cut ties with the Cambridgeshire company after a 1997 television programme showed workers mistreating animals. GSK has since re-employed the firm to conduct animal toxicology tests.

Members of Pro-Test, a Cambridge-based group that supports animal testing, purchased ten shares of GSK in solidarity.

Critics pounce on bubble-fusion blunder

A group of researchers making high-profile claims about fusion energy has admitted to accidentally using equipment different from that reported in their most recent paper.

In January, Rusi Taleyarkhan and his team published the latest in a series of papers that claimed to see neutrons characteristic of fusion reactions coming from collapsing bubbles in organic fluids. Such 'bubble fusion' has been hailed as a source of cheap, renewable energy, but there have been serious questions about the work's validity (see www.nature.com/news/2006/060306/full/060306-1.html).

The paper says that four independent detection methods were used, one of which was a boron trifluoride gas proportional tube. The erratum, published by Taleyarkhan and his colleagues in *Physical Review Letters* on 2 May, reveals that this part of the apparatus was actually a lithium iodide crystal scintillation detector. Critics interpret the admission as a sign that the group's fusion claims are further unravelling.

Taleyarkhan, who is based at Purdue University in West Lafayette, Indiana, says the equipment error does not affect the data, analysis or conclusions. A university review of the work is due by 1 June.

Disgraced cloner accused of embezzling grants

Woo Suk Hwang, the once-fêted cloning researcher whose career collapsed amid charges of fraud, was indicted in South Korea on three charges last week.

On 12 May, the Seoul Central District prosecutor's office charged Hwang with embezzling 2.8 billion won (US\$3 million). He is said to have committed fraud by knowingly using fabricated data to apply for research funds.

Hwang was also charged with violating a bioethics law that bans the purchase of eggs for research. Five other researchers on his team were charged with various offences. The cases could now proceed to trial; Hwang's lawyer says that he is preparing a statement.

LOOK AND LEARN

Prevailing wisdom says the adult brain cannot learn to see if it had no visual stimulation during childhood, but blind people in India seem to be breaking all the rules. **Apoorva Mandavilli** reports.

Doctors gave SK his first pair of glasses in July 2004. He had been too poor to afford a pair before — but then he was a 29-year-old blind man, what use were glasses to him? Had he been given glasses as a child they might have helped him overcome his congenital aphakia — an extremely rare condition in which the eyeball develops without a lens. Yet his chances of being diagnosed, let alone treated, in the poor Indian village in which he was born were slim. As a result, SK was living in a ‘hostel for the blind’ with no running water when the doctors arrived from New Delhi.

SK's doctors weren't sure how much sight he would gain, or if he would comprehend what he saw. For the first year, he had only the most basic visual skills. He could recognize simple two-dimensional objects but anything three-dimensional, even an everyday object such as a ball, was beyond him. All this was consistent with the idea of a ‘critical period’ in vision: that if you haven't learned to see by a certain age, you never will.

But 18 months after getting his glasses, SK surprised everyone. He had begun to make sense of his world, building his visual vocabulary through experience and recognizing more complex objects with varying colours and brightness. In doing so, he turned one of the most fundamental concepts in neuroscience on its head.

“Twenty-nine years without any normal vision? I would have said that's a life sentence,” says Ron Kalil, a visual neuroscientist at the University of Wisconsin in Madison. For Kalil and other experts, the impossible now seems possible. And while the scientists might be amazed by the brain's adaptability, the real winners are the countless blind people — both children and adults — who had been considered untreatable.

Light work

SK is the first success of Project Prakash (Sanskrit for ‘light’) launched in 2003 and run by Pawan Sinha, a neuroscientist at the Massachusetts Institute of Technology. Originally a humanitarian effort to help blind children in India, under Sinha's guidance Project Prakash has blossomed into a chance to investigate how we learn to see.

The idea of a critical period for vision grew out of research in animals. In the 1970s, Torsten Wiesel and David Hubel at Harvard Medical School famously shut one eye of a week-old kitten. They found that closing the eye for only a few weeks caused the kitten's



After 29 years of being officially blind, SK is learning to see — and defying neuroscience in the process.

visual cortex, the part of the brain that deals with sight, to develop abnormally. But similar experiments had little effect on adult cats.

After further work in monkeys, Wiesel and Hubel suggested that there is a critical period during the first few months of life when normal vision must develop. In 1981, they shared the Nobel Prize in Physiology or Medicine for this and other work.

“The work in India challenges the clinical notion that treating a blind child beyond the age of eight is hopeless.”

The notion of a fixed critical period has been crumbling steadily for the past few decades, and the brain is now seen as much more flexible — able to grow new nerve cells and to adapt long after childhood. For instance, Uri Polat and his colleagues at Tel Aviv University in Israel showed in 2004 that adults with a lazy eye can be trained to improve their vision¹ — contrary to conventional expectations.

In the strictest sense, SK's case does not con-

tradict Wiesel and Hubel's finding. Without glasses, his visual acuity was 20/900, far worse than the standard 20/20 for normal vision or even the World Health Organization's definition of legal blindness at 20/400. Like others with aphakia, SK could sense light and movement, which arguably allowed his visual cortex to develop normally.

With his glasses, SK's acuity jumped to 20/120, although he still saw objects as separate regions of colour and brightness, and struggled to put them together as a whole. He saw a cow, for example, as patches of black and white, each a separate object, until the cow moved. As soon as there was movement, SK was able to recognize that the set of objects made up a cow. Over the next 18 months, his acuity remained stuck at 20/120, but he was able to stop relying on motion to integrate objects and learned to recognize them even when they were still.

Critical point

Under conventional dogma, SK shouldn't be able to do this, but where scientists went wrong, says Sinha, is in applying the notion of critical period too broadly. They assumed, incorrectly, that if the eye's vision doesn't improve, neither does the ability to interpret what you see. “What is so exciting about Sinha's work is that he is showing that's not the case,” says Lynne Kiorpes, a neuroscientist at New York University. “People can learn to use the vision they have.”

As well as being Sinha's native home, India is a natural choice for studying blind children. One-third of the world's 45 million blind people live in India, according to the non-profit organization Orbis International. Caught early, half of the cases of childhood blindness could be treated or prevented — but they aren't.

Some parents see blindness as punishment for sins in a previous life. Most can't afford the treatment. Born poor, and neglected in the broader struggle for survival, more than 60% of blind children in India die before reaching adulthood, according to Orbis. Those who survive tend to be shipped off to special schools or hostels, where they resign themselves to making candles or weaving baskets.

“The first thing that prompted me was seeing these numbers, the humanitarian goal was just so evident,” says Sinha. After touring villages in the summer of 2002, Sinha joined forces with Dr Shroff's Charitable Eye Hospital on the outskirts of New Delhi. By Western standards the hospital is unassuming, but it is among the best equipped in the country.

With funding from international organizations, volunteers from Dr Shroff's hospital

head into under-served communities and screen hundreds of children for sight problems. In many cases, the children are beyond help. But if the condition seems reversible, the hospital offers to treat them, often for free. Many suffer from cataracts and, with only minor interventions, could well have normal vision.

With Project Prakash, volunteers from Dr Shroff's hospital have begun visiting schools for the blind for the first time. The project aims to treat 15 'hopeless' cases from these schools each year, and to enrol them in further experiments. SK is the oldest by far: most are aged 7 to 11. Sinha plans to measure the children's visual abilities before and after treatment to learn how much vision is needed for recovery to be possible.

Alternative view

SK's case has already provided new ideas about how vision develops. Studies on individuals who recover sight late in life are rare, and researchers who have measured acuity found that, as with SK, it did not improve over time². Those scientists did not ask, as Sinha did, whether other visual skills might eventually improve.

"If they'd looked at it that way, then they probably would have found all sorts of things they weren't looking for," says Nigel Daw, a neurobiologist at Yale University. "This will lead people to do more experiments along those lines. It's ground-breaking work in that sense."

Sinha's work suggests that different visual abilities might have different critical periods. For example, detection of motion is one of the first abilities to develop, perhaps even hard-wired at birth; understanding of colour and full stereo vision develop later, and visual integration — the process that allows SK to resolve separate objects into a cow — might take even



Pawan Sinha (left) gauges the progress of an Indian child who gained his sight later in life than usual.

longer. "This project allows us to refine the very broad and vague idea of a critical period," Sinha says.

And it challenges the clinical notion that treating a blind child beyond the age of around eight is hopeless. Most children with reversible causes of blindness can sense light, and so may be able to recover. In fact, says Sinha, the complete blindness created in animal experiments is rarely, if ever, found.

Wiesel himself maintains that a critical period still exists, but that scientists should be more open-minded about how they interpret it for clinical use. "There is a critical period and we should try to help kids as soon as possible," Wiesel says. "But if for some reason that's not possible, and there's some vision left, certainly recovery is possible."

SK's case has led Sinha to some interesting tangents. For example, children with autism

also have trouble integrating different parts of an object into a whole — they literally can't see the forest for the trees. It turns out that these children also have trouble with motion perception^{3,4}, which bolsters Sinha's hypothesis. He is now studying visual skills in autistic children.

For Project Prakash, this is just the start. Soon after the project began visiting schools for the blind, Dr Shroff's hospital petitioned the Indian Supreme Court, demanding that at least those children who can be helped are given a chance. As a result, India's Supreme Court has ruled that before being admitted to a blind school, every child must be examined by an ophthalmologist.

On the research front, Sinha plans to use imaging techniques — primarily functional magnetic resonance imaging — to study the brains of children and adults whose sight is restored later in life. He wants to see how much of the visual cortex in these people can be stimulated by light. Sinha also hopes to discover how much of the cortex has been taken over by other functions such as hearing or touch, and whether this change is reversible. "I think that is a very, very fundamental question," he says.

Presumably, the children might also recover different skills to varying degrees. Such information might help doctors design rehabilitation programmes and may offer clues to which visual abilities come prewired at birth and which develop over time. "Slowly but surely, the evidence is coming forward to indicate that the brain is much more adaptable than we suspected," says Kalil. "Work such as Pawan's should tell us to go looking for more."

Apoorva Mandavilli is senior news editor for *Nature Medicine*.

1. Polat, U., Ma-Naim, T., Belkin, M. & Sagi, D. *Proc. Natl Acad. Sci. USA* **101**, 6692–6697 (2004).
2. Fine, I. et al. *Nature Neurosci.* **6**, 915–916 (2003).
3. Bertone, A., Mottion, L., Jelenic, P. & Faubert, J. J. *Cogn. Neurosci.* **15**, 218–225 (2003).
4. Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L. & Stone, W. L. *Psychol. Sci.* **14**, 151–157 (2003).



Staff at Dr Shroff's Charitable Eye Hospital are trying to reduce the levels of blindness in India.

BATTERIES NOT INCLUDED

Among their many talents, bacteria are the world's best electrochemists, creating a life-powering flow of electrons in a startling range of conditions. In the first of two features, **Nick Lane** asks what limits, if any, constrain this ability. In the second, **Charlotte Schubert** meets the people trying to put this microbial ingenuity to practical use.

What can't bacteria do?

In 1977 Engelbert Broda, a physical chemist, then at the University of Vienna in Austria, made a startling prediction¹. Nitrate and ammonia can and will react together in the way that oxygen and organic carbon molecules do. And so, in theory, it would be possible for a creature to make a living by breathing nitrate and eating ammonia, just as we do by breathing oxygen and eating proteins, carbohydrates or fats. The creatures Broda had in mind for this bizarre diet were not some race of aliens floating in a soup of ammonia under red, nitrate-laden skies, but earthly bacteria of a sort never before imagined.

The discovery of bacteria capable of this 'anaerobic ammonia oxidation' or anammox reaction, in the late 1990s², was a high-water mark for the rising tide of what might be called microbial triumphalism. The anammox reaction, it transpires, is a dazzlingly hard trick to pull off — it produces hydrazine (rocket fuel) as an intermediate, which the bacteria have to tuck away in internal sacs made of lipids.

If bacteria have found a way to take advantage of nitrates' propensity to oxidize ammonia, despite having to cope with a toxic by-product, is there anything they cannot do? Subsequent discoveries — of bacteria that 'breathe' metal oxides, or feed on bleach — have strengthened the case for bacterial omnipotence. Last month, one of the last gaps in the bacterial skill set was closed when some of them (in concert with helpful members of another class of single-celled organism, the archaea) were discovered using nitrates to get energy from methane³.

And yet there are still some energy-producing pathways down which no bacteria have been seen to stroll. No known bacteria seem to live by oxidizing manganese; none derive their energy from molecular nitrogen; and no photosynthetic bacterium is known that uses ammonia, phosphine (phosphorus hydride), or even methane, although the laws of thermodynamics show that all those reactions could provide energy. Are such organ-

isms really absent, or have we just not found them yet? And if they are absent, does this reflect a fundamental limitation on life's ingenuity? Or does it reveal set truths about the Earth's environmental chemistry?

"I can imagine that evolution's random walk has failed to locate some reactions," argues Franklin Harold, an emeritus professor of biochemistry and molecular biology at Colorado State University in Fort Collins. "If no bacteria live by molybdenum reduction, no profound principle is indicated." But there is some reason to believe that the missing tricks could explain why bacteria specialize so narrowly, forming rich metabolic networks and ecosystems only through collaboration. Indeed, bacteria's inabilities may be fundamental to the question of why there are ecosystems at all, rather than one all-conquering bug.

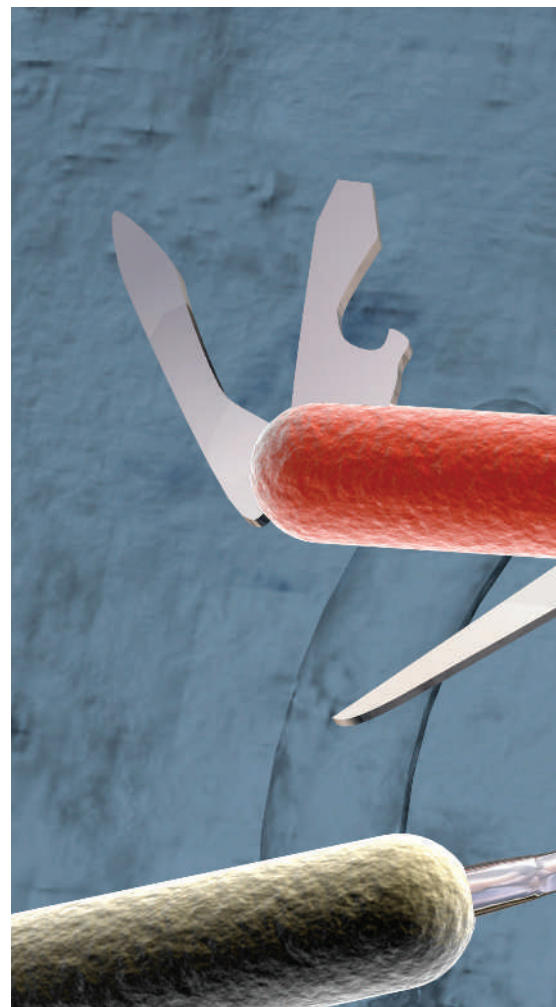
Working in pairs

"It's a dangerous game to put up your hand and say 'bacteria can't do this or that', as we probably only know about 1 to 5% of bacteria on the planet" says microbiologist Ken Nealson, at the University of Southern California, Los Angeles. "But I'd wager that there are constraints on what's possible, beyond whether a reaction yields energy; some of these are still an incredible, wonderful mystery."

The common factor in all the forms of respiration known is the involvement of what chemists call a redox pair — an electron donor and an electron receptor. The process by which electrons are stripped from the donor is called oxidation, because oxygen is the most familiar stripper of electrons (and also a very powerful one). The electrons are transferred to an acceptor, which is said to be reduced; hence the term 'redox'. Non-chemical ways of generating energy are conceivable (see 'The first electrochemist') but don't seem to be used.

Life's dependence on redox reactions led the great geneticist Hermann Muller to describe it as "a fancy form of rust". But it is rust with a

purpose. As the electrons go from willing donor to willing acceptor, they go down a thermodynamic gradient; this means they lose energy. In respiring cells, the energy from this flow of electrons is used to build up a gradient of protons across membranes. When the protons flow down this artificially created gradient back across the membrane, ATP is created



— a molecule that stores energy in a form that can be given up wherever it is needed to drive the cell's molecular machinery. (Humans may prefer to use these electrons for other purposes; see page 277.)

In the case of aerobic respiration, used by most animals and many bacteria, the electron acceptor is oxygen, but it can be anything that has a hunger for electrons: carbon dioxide, nitrate, perchlorate (used in bleaches) or plain protons will do. All that matters is that the recipient's redox potential — its tendency to acquire electrons — is higher than that of the donor. The greater the difference in potentials, the more energy is released (see graphic, overleaf).

In photosynthesis, the energy of the Sun is used to tear electrons from unwilling donors, driving redox reactions thermodynamically uphill. The absorption of energy by a chlorophyll molecule turns it into a fierce oxidant. Like a dentist pulling teeth it then wrenches electrons from a bystander that would normally be loath to surrender them. The electrons are ultimately transferred to carbon dioxide (the electron acceptor) to form sugars.

Whether the electrons come from willing donors, as in respiration, or unwilling ones, as in photosynthesis, the bacteria need a way to get rid of the reduced molecules where the electrons accumulate once the reactions have

run their course. In anaerobic systems, the reduced molecule in question will often be hydrogen. "Unless the hydrogen is removed, the whole ecosystem gums up; that's why hydrogen consumers such as methanogens [archaea that produce methane] are so important," says Nealson. Other organisms in the ecosystem benefit: "By removing hydrogen, shifting the equilibrium, they make other processes thermodynamically possible, such as using protons as electron acceptors, which generates more hydrogen."

Current thinking

A build up of end products is not the only obstacle. The need for specially lined rocket-fuel tanks that the anammox reaction forces on bacteria highlights the fact that intermediates can also be a problem. So do recent findings about bacteria that oxidize perchlorate, notes Stuart Ferguson, a biochemist of the University of Oxford, UK.

So far, six bacterial species are known to use perchlorate as an electron acceptor⁴. At first, the key component was thought to be an enzyme shared by these species that normally passes electrons to nitrate or nitrite, and which serendipitously happens to be able to deal with perchlorate too (the receptor molecules are not dissimilar in size and shape). But more



M. ZHU

"To make a certain reaction work, a bacterium has to occupy a particular niche."

— Robert Blankenship

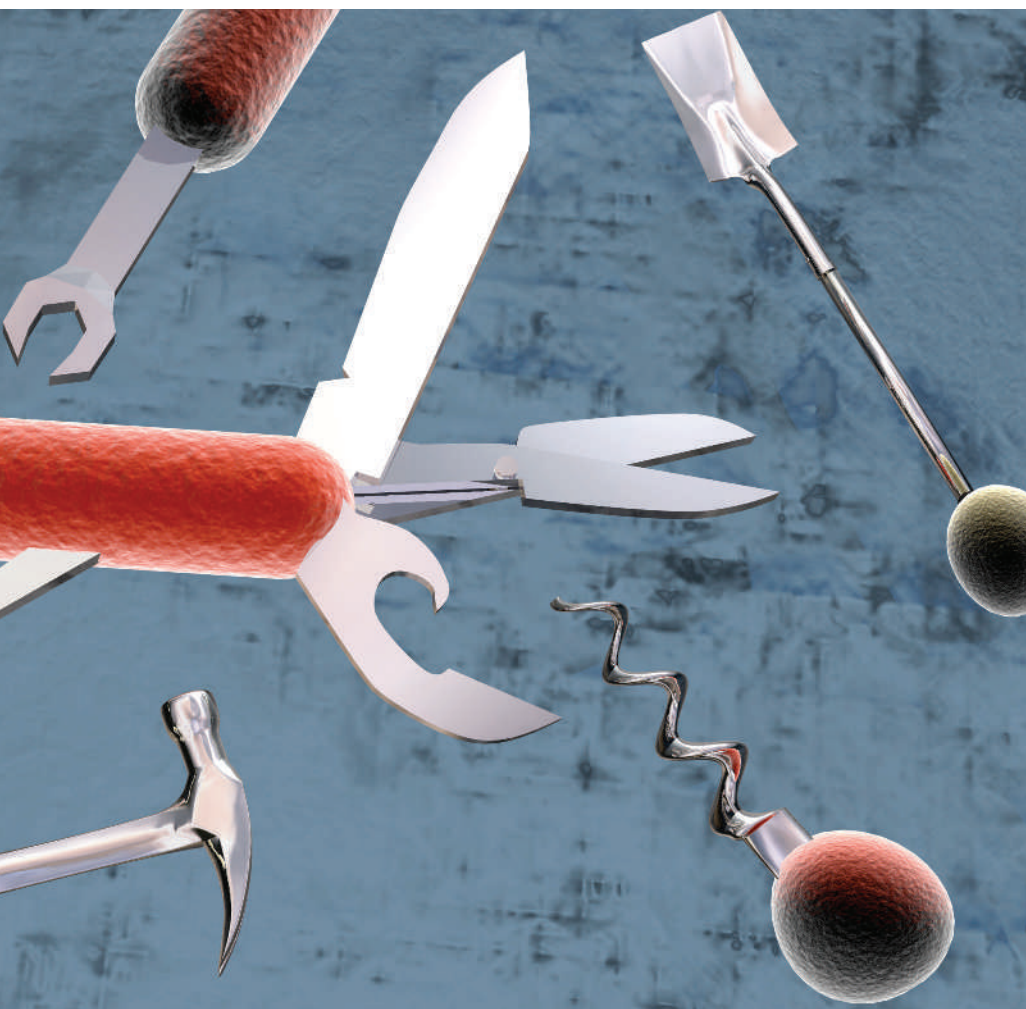
recent work has shown that the crucial common denominator is in fact an enzyme that can detoxify the most toxic intermediate, chlorite⁵. Bacteria with the enzyme chlorite dismutase can cope; those lacking it could not. Being able to cope with toxicity, not the ability to carry out the appropriate redox chemistry, is the bottleneck.

Is this a general truth? "Toxicity can be a big challenge for microbes," says geochemist Donald Canfield of the University of Southern Denmark. Canfield thinks that there may be molecules that are too toxic for bacteria to deal with. "I don't know of any microbes that can oxidize phosphine, for example. Perhaps it's just too toxic — it ignites spontaneously in air." This is despite there being environments such as swamps and rice paddies where phosphine is plentiful. "From a thermodynamic point of view it would make good sense," argues Canfield. In oxygen-free settings, phosphine's spontaneously combustible nature might be less of a problem. If there are phosphine-eaters, the most likely place for them to turn up would be in anaerobic sediments such as those of Lake Taihu in China — a site of ongoing research into the phosphate cycle.

Another factor that might stump bacteria is kinetics — the speed at which a reaction takes place naturally. For example, a potential source of energy in many marine sediments is the reaction of iron oxides such as rust with hydrogen sulphide that can bubble up from below. "But this reaction happens so fast naturally that, apparently, no bacteria can tap into it for energy," says Canfield. "They would have to come up with an enzyme that could actually slow it down."

On the other hand, some thermodynamically favourable reactions are hard to get started, and go slowly. Enzymes are good at speeding up such reactions through catalysis, but the degradation of certain compounds needs a huge pulse of energy to get going. For such compounds, with a very high activation energy, the reaction is too slow to be profitable for microbes despite their enzymatic cleverness.

"A good rule of thumb is to look for what



accumulates", says biochemist Rolf Thauer, at the Max Planck Institute for Terrestrial Microbiology in Marburg, Germany. Lignin is a good example. When lignin, a key component of hard wood, first evolved around 375 million years ago, nothing could metabolize it; the great reserves of coal in the Carboniferous period attest to that failure. Even today, it remains a challenge, and one that is met for the most part by aerobic fungi, says Thauer. "The molecule is too large for most bacteria to handle, and its activation energy is too high."

Conversely, a lack of build up can be evidence of microbial activity. Geochemists knew that methane was 'missing' from ocean sediments long before the anaerobic bacteria and archaea (another group of microbes) that oxidize it were discovered. It should have been accumulating, but wasn't, so something had to be oxidizing it. "Microbiologists said 'that's not possible,'" recalls Thauer, "but clearly it was going somewhere." Benzene is a similar case. Its disappearance from anaerobic environments means something must be consuming it. But the organisms and metabolic pathways involved have yet to be elucidated.

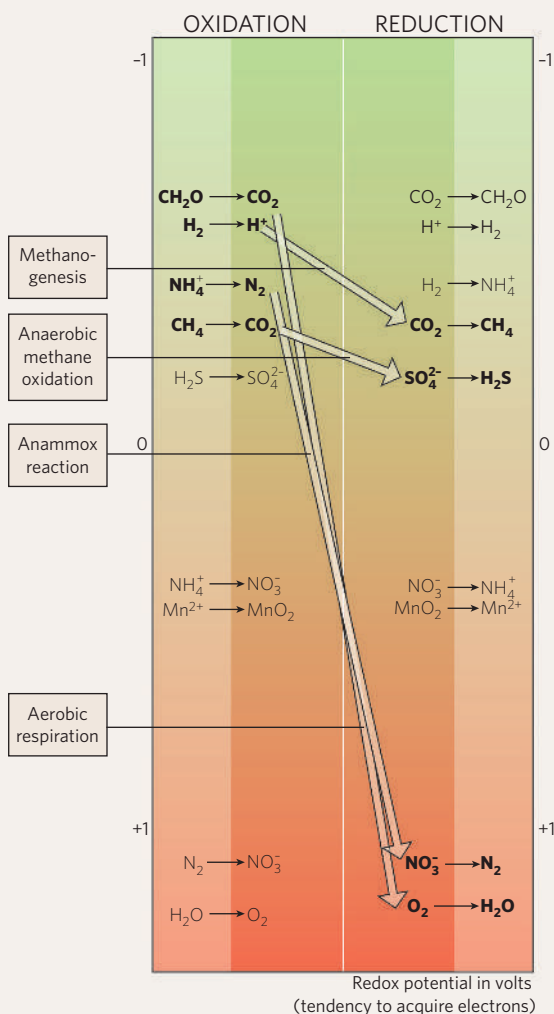
Accumulating evidence

Perhaps the most interesting accumulation is of the molecule currently filling 80% of your lungs. Atmospheric nitrogen is mostly made by bacteria that reduce nitrates. Reducing the nitrogen further, to make ammonia, could in principle be a way of getting energy out of various organic fuels; but this doesn't seem practical. "Bacteria do reduce nitrogen," says Friedrich Widdel, at the Max Planck Institute for Marine Microbiology in Bremen, Germany. "But none can glean energy from this process: the activation energy needed to break the triple bond of nitrogen is too high."

Widdel points out that the opposite process, using nitrogen as an electron donor and thus forming nitrate, is also possible. He thinks that ecological constraints may be blocking

MIX AND MATCH REACTIONS

Bacteria and archaea can tap into the energy made available when electrons released from an oxidation reaction are used in an electron-absorbing reduction that is lower down the energy scale. (The length of the thick arrows indicates the amount of energy released.)



the use of this reaction by bacteria. "Perhaps the problem is the high energy 'hills' on the way to nitrate," he speculates. "But photosynthetic organisms should have unlimited energy: they 'ought' to be able to use nitrogen as an electron donor."

Robert Blankenship, a biochemist at Arizona State University in Tempe, agrees that

there are many potential forms of photosynthesis that bacteria don't seem to use, and also suggests an ecological explanation — scarce resources. "Because sulphur compounds are used as electron donors for photosynthesis, you might predict that phosphorus and nitrogen compounds would be used too, but to the best of our knowledge they're not," he says. "I suspect they're too important for other purposes — they're in short supply, and limit growth because they're needed for proteins and nucleotides."

Creatures of habit

If there were no other electron donor available, this constraint might not be so strong. But there almost always is. Cyanobacteria, algae and plants benefit from a type of photosynthesis that uses water as its electron source; this process can take place anywhere where there is sunshine, water and carbon dioxide (the electron acceptor).

What's more, with this kind of oxygen-generating photosynthesis, the reduced compounds that might otherwise be used as electron donors in photosynthesis are mopped up: the compounds react chemically with oxygen in the surroundings, or properly equipped bacteria burn them up in respiration. Few reduced compounds survive long where light is plentiful, because of the oxygenating effects of photosynthesis. This may well be why ammonia-based photosynthesis, another possible form of metabolism that Broda discussed in the 1977 paper¹, has never been seen.

These constraints may hamper the iron-oxidizing photosynthetic bacteria first discovered by Widdel and his colleagues in the early 1990s⁶. The purple bacteria use soluble ferrous iron as an electron donor for photosynthesis, precipitating insoluble ferric iron as rust. But they seem to be restricted to sediments rather than open waters. Part of the problem is that ferrous iron is rarely found high in the water column, where it is quickly oxidized; so iron-oxidizing bacteria thrive only in

THE FIRST ELECTROCHEMIST

Bacteria may be the best electrochemists going, but their obsession with redox reactions is a curiosity in itself. Why not use heat or mechanical forces for energy? Plenty of bacteria live in hydrothermal conditions — life on Earth may even have originated there. Yet none has been found that can tap the heat gradient of vents directly

— perhaps by using a long thin cellular structure to stretch across the gradient from cold to hot.

Or why not use water currents? Bacteria often move around by rotating rigid corkscrew-like flagellae, using tiny molecular motors powered by redox chemistry. There is no reason in principle not to fix the

cell to a base, use water currents to rotate the flagellum in reverse and capture the mechanical force.

The dependence of all microbes on electrochemistry tells us something about early life on Earth. At the very least, it shows that the last ancestor of all microbes must have been an electrochemist. Either it was so

successful that it supplanted all life using other forms of energy in its first flush of youth, or else the origin of life itself involved redox chemistry. "We may never know which until we see how life operates on other planets," says microbiologist Ken Nealson, at the University of Southern California, Los Angeles.

N.L.

the sediments of shallow waters where mud low in oxygen is exposed to sunlight.

The things bacteria don't do mostly seem to be things the environment doesn't easily let them do. But although, as a group, bacteria can make use of more or less every reaction within thermodynamic and ecological limits, any given bacterium can probably only manage a few of them. Given the apparent ease with which bacteria and archaea swap genes, this might seem surprising. Why don't photosynthetic sulphur bacteria, for example, ever reduce sulphate in the dark? Why are different bacteria needed to oxidize ammonia to nitrate and then reduce the nitrate back to nitrogen. In short, why is there not one super-bacterium that does everything?

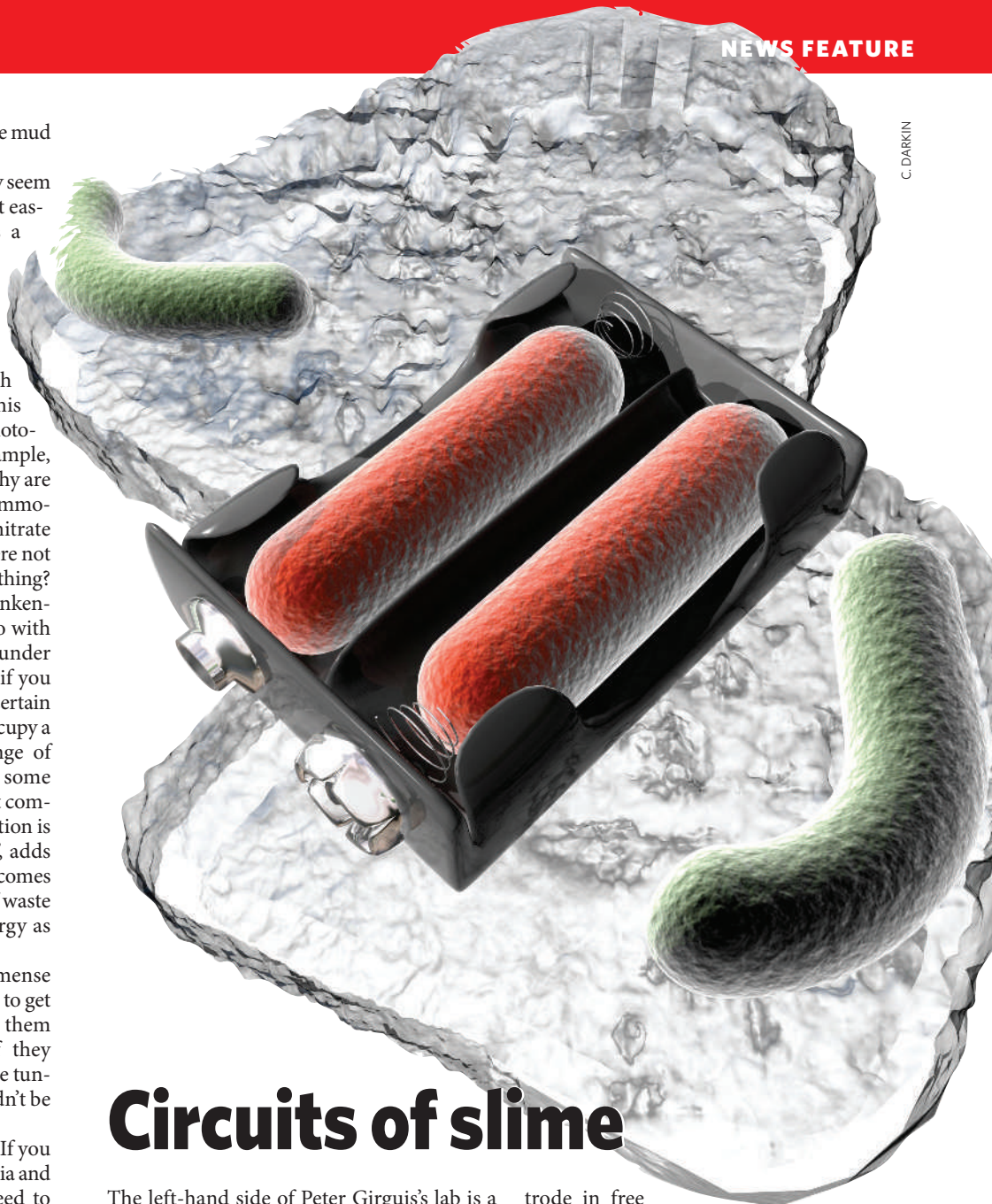
"That's a vexed question!" says Blankenship. "I'd say it was something to do with redox reactions running one way under some conditions, and another way if you change the conditions. So to make a certain reaction work, a bacterium has to occupy a particular niche." If a specific range of redox potentials are needed for some processes to work, but others are not compatible with that, then "the best solution is to cooperate with your neighbour", adds Blankenship. The ecosystem thus becomes a way of dealing with the build up of waste products, and getting as much energy as possible out of the environment.

If Blankenship is right, the immense subtlety with which bacteria are able to get one thing done may be what stops them from getting everything done. If they weren't constrained by their own fine tuning of redox conditions, there wouldn't be any diverse microbial systems.

And there's an obvious corollary. If you really want to understand how bacteria and archaea make their livings, you need to study communities, not individuals, whether they be tightly coupled partnerships such as those between archaea and bacteria that oxidize methane or the looser, more complex communities found wherever the redox potential of a sediment varies with depth. And that approach requires, among other things, a new culture of microbiology. As Neelson puts it: "When I was a student, you were thrown out for working with mixed cultures, but today the interest is in how bacterial consortia operate. If you fish out two bacteria in a cling, they're doing something together, but we've hardly begun to look."

Nick Lane is a science writer based in London.

1. Broda, E. Z. *Allg. Mikrobiol.* **17**, 491–493 (1977).
2. M. Strous et al. *Nature* **400**, 446–449 (1999).
3. Raghoebarsing, A. A. et al. *Nature* **440**, 918–921 (2006).
4. Coates, J. D. et al. *Appl. Environ. Microbiol.* **65**, 5234–5241 (1999).
5. Xu, J. & Logan, B. E. *J. Microbiol. Meth.* **54**, 239–247 (2003).
6. Widdel, F. et al. *Nature* **362**, 834–836 (1993).



C. DARKIN

Circuits of slime

The left-hand side of Peter Girguis's lab is a mucky place. Dried mud flecks benches, beakers and scales, and buckets of slime clutter the floor. Fresh ooze is flown in regularly from California. A skein of pipes fed by a 2,000-gallon tank of sea water in the basement keeps everything moist. On the floor, wires gush out of dismembered plastic cocoons, and electrodes poke out of various tanks.

The messy workplace is a design shop for electrical generators. This summer, microbiologist Girguis and his colleagues at Harvard University in Cambridge, Massachusetts, will shove graphite electrodes, like those used in this lab, into the sea floor of Monterey Bay in California. The current that flows between the electrodes will power a suite of scientific instruments, including a wave and tide meter, and a fluorometer that measures levels of chlorophyll in sea water.

Girguis's work depends on the fact that when bacteria respire they pull electrons off organic debris. Catch those electrons on an electrode that is hooked up to a second elec-

trode in free water, say, or in air, or in another layer of sediment where the microbes use a different sort of electrochemistry (see page 274) — and the electrons will flow from the first to the second electrode. That gives you a current with which you can power things.

The currents produced by these microbial fuel cells are small, but their potential, in the non-voltage sense, may be surprisingly large. After all, microbes can produce electrons from sediment, sewage, food scraps or pig slop. Girguis, a self-confessed gearhead, takes some delight in a simple demonstration of their current-generating abilities using just a bucket, some hardware-store supplies and a bag of cow manure.

People have been trying to harness the power of microbes in this way since the early 1900s, although with little practical success. Engineering advances and molecular biology's new tools, some of which have a home in the clean half of Girguis's lab, have allowed him and like-minded

researchers to breathe life into the field.

"This technology would transform the world if it ever became reliable and could be used on a large scale," says Bruce Rittman, director of the Center for Environmental Biotechnology at Arizona State University's Biodesign Institute in Tempe. But he admits that is still a dream. "We are quite some distance from capturing the benefits at an economic scale." To get there, he estimates that researchers will need to produce fuel cells with power-generating abilities that are a hundred times greater than those of today's cells.

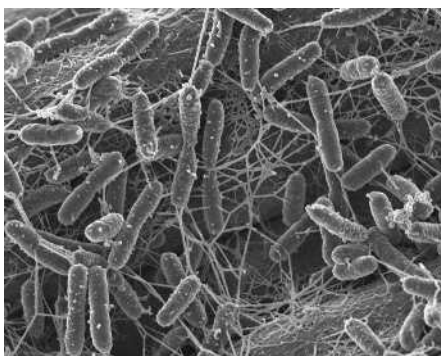
"All of microbial fuel-cell research invariably has to deal with the interface between microbes and the electrodes," says Girguis. One approach, favoured by engineers, is to try to improve the cell's design and the materials from which the electrodes are made. Biologists, however, aspire to better understand the microbes themselves, with the aim of informing more sophisticated designs — or designing better bacteria.

Natural lighting

Fuel-cell builders have been adding electron mediators — small molecules that help electrons move around — to their systems for years to increase the current. But another way to enhance performance may be to increase the physical contact between the bacteria and the electrodes.

Last year, Derek Lovley at the University of Massachusetts, Amherst, revealed that members of a group of bacteria called *Geobacter* extrude protein spines, or pili, which conduct electricity². And they are not the only ones, according to findings presented this February at a conference sponsored by the US Department of Energy, and held in Bethesda, Maryland.

Yuri Gorby from the Pacific Northwest National Laboratory in Richland, Washington, reported that nanowires sprout from a range of bacteria, including the iron-breathing microbes *Shewanella* and the photosynthetic bacteria *Synechocystis*. Using *Synechocystis*, he and his colleagues have made a solar-powered form of microbial fuel cell. The key to the nanowires' conductivity, Gorby and his colleagues found, was the presence of certain cytochromes —



Multi-talented: microbes could help remove toxins from waste water, and power the treatment plants.

electron-shuttling molecules found in bacterial membranes. Bacteria lacking the nanowires or cytochromes are poor conductors of electricity.

Researchers are already thinking about how to harness the findings. Gorby's colleague, microbiologist Ken Nealson at the University of Southern California in Los Angeles, points out that *Geobacter* and *Shewanella* have

"This technology would transform the world if it ever became reliable and could be used on a large scale."

— Bruce Rittman

evolved to transfer electrons to scraps of metal found in nature, not to an artificial electrode. Encouraging them to evolve electrode compatibility might yield a great deal of improvement. Nealson and Lovley say that 'electrode-adapted' bacteria could one day produce fuel cells with at least 100-fold greater power-generating abilities compared with current cells.

Such 'improved' bacteria might be too specialized and delicate to thrive in marine sediments. But they could be put to use in purpose-built closed bioreactor systems, where they could generate electricity by chewing up biomass. "It's the best approach we have to developing an industrial application or producing a lot of power," says Girguis.

While the biologists extol the potential of nanowires, Bruce Logan, an environmental

engineer at the Pennsylvania State University in University Park, prefers to concentrate on the chemistry and physics of the fuel cells. Logan says he hasn't seen much practical advance from understanding the biology of the microbes. "It's not clear whether we are limited by the bacteria or the physics of the system," he says. And although he is not dismissive of the microbiologists — indeed he has just started collaborating with Nealson — he argues that "most of our improvements have been geared towards the physics and chemistry, suggesting that that is the major roadblock."

Energy on the cheap

Logan works with industrial runoff, animal waste and sewage. At least 1.5% of US electricity production goes into wastewater treatment, he says. He wants to change that equation, by developing systems that produce electricity as well as cleaning the water.

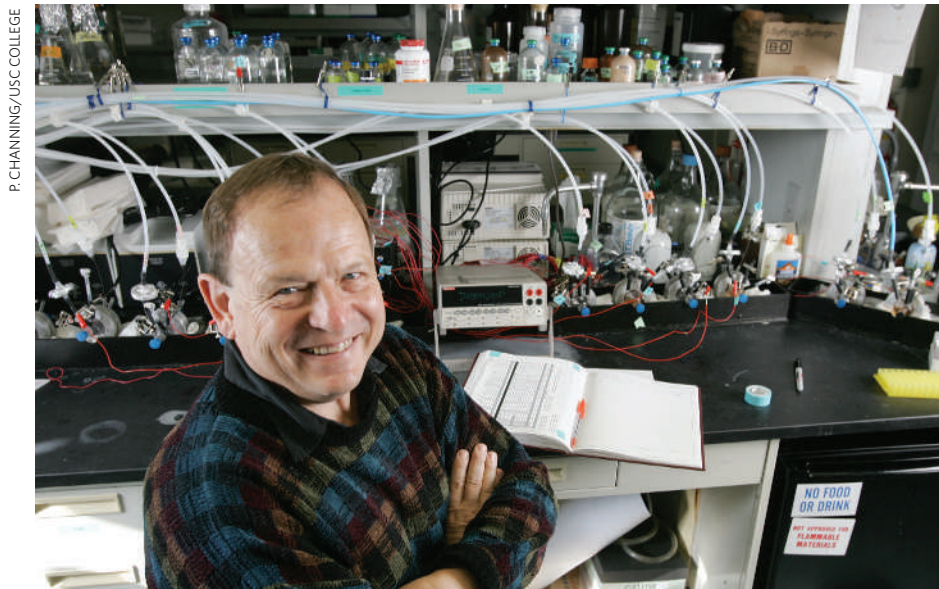
The field of wastewater treatment by fuel cells was opened up by Byung-Hong Kim, a microbiologist from the Korea Institute of Science and Technology in Seoul; Kim developed a system that didn't need added electron mediators in the 1990s³. Since then engineers, such as Logan, have worked on various ways of improving the design and the power output of these cells. Logan's most recent table-top device yields the equivalent of 15.5 watts per cubic metre of household waste water flowing through it⁴.

By the standards of a conventional fuel cell, such as one powered by hydrogen, that's a very poor yield. But with scale-up, Logan estimates that the waste water from 100,000 people, or a large industrial plant, could produce 0.8 megawatts, enough to power about 500 homes. Such a system would also scrub waste water of troublesome compounds such as ammonia, which is consumed by the microbes, and produce 50–90% less sludge than conventional methods.

Unfortunately, the electron-gathering anodes in Logan's laboratory systems are made from moderately expensive graphite or carbon paper, and the cathodes need to be coated with very costly platinum for the best effects. Willy Verstraete, an environmental engineer at Ghent University in Belgium, says that, extrapolating



The power of mud: soil bacteria free up electrons as they respire. When used to induce a current between electrodes, these can light up a lamp.



Ken Neilson thinks technologies using microbes could be improved by knowing about their biology.

linearly from the costs of the prototypes, the capital expenditure for a full-scale plant would be five to ten times higher than it is for today's treatment plants. So the hunt is on for cheaper materials.

Logan points out that the technology does not have to be cheap, just cheaper than the current techniques. The United States already spends about \$25 billion each year on domestic wastewater treatment, he says. He is working on a design of fuel cell that he says could be cost effective and scalable; he hopes this will yield ten times as much power per cubic metre of waste water as current designs do.

Power to the people

Verstraete is more cautious. "Industry is expressing a lot of interest," he says, "but they are not putting the money on the table." He says that companies want to see yields as high as 1,000 watts per cubic metre — and to see them in prototypes a thousand times larger than the milk-carton-sized devices now honing the technology. He sees a role for the technology in cleaning certain nitrogenous pollutants out of waste water, or chewing up effluents from the digesters that convert biomass to methane on some farms. But it's still a long road to the first microbial car (see 'And what if it worked?').

Meanwhile, the powering of marine instruments seems to be the application moving the fastest. Girguis's work on Monterey Bay sediments follows the lead of pioneers such as Lenny Tender at the US Naval Research Laboratory in Washington DC. Tender was one of the team that first successfully extracted current from marine sediment more than five years ago⁵. This summer, Tender plans to launch a microbial fuel cell that powers an acoustic device for measuring water velocity. Taking measurements every 4 hours it will consume an average of just 0.1 watts.

Environmental monitoring on the land and

the ocean surface is increasingly powered by solar cells that can work for years unsupervised. The ocean floors, being abysmally short of sunlight, need a different approach, and Tender, Girguis and their colleagues think that microbial-powered electricity is the solution. This would obviate the need for battery replacement that currently drives up the costs of such studies.

"They have a lot more work to do before they can call this a real proven technology," says James Bellingham, chief technologist at the Monterey Bay Aquarium Research Institute, who is working with Girguis. "But these guys have come very far very fast."

Which other niche areas might be successful is less clear. For his terrestrial applications, Girguis is collaborating with an architect who specializes in new materials. Together, they plan to design LED lighting systems that can

be powered by a bucket of food scraps or animal waste. He says such devices could be mass-produced for about ten dollars each and could help generate light for many of the 1.6 billion people living outside the world's electrical grid. Although such people are obvious candidates for solar power, Girguis says the microbial option has the advantage of being much less expensive to manufacture and easier to operate and repair. And it can work in the dark.

The technology need not always be cheap and cheerful. Rittman has a grant of \$100,000 from NASA to design a compact microbial fuel cell that consumes human waste during a manned astronaut mission. But like many researchers in the area, he dreams of bigger opportunities on Earth. If microbial fuel cells could be adapted for the consumption of plant biomass, they could in theory produce twice the amount of electrical energy that combustion-driven generators can, says Rittman. Again, however, the best that has been done is not quite that good, and is stuck on the lab bench for now.

But Neilson, for one, is confident things can improve, especially if links between different disciplines housed in different departments can be strengthened. "Our job here is to understand how the bacteria work, and once we know that we can get together with the engineers. Then you might find out how to make the ideal fuel cell." He laughs, "maybe it will drive itself across campus."

Charlotte Schubert is senior News and Views editor for *Nature Medicine*

1. Potter, M. C. *Proc. R. Soc. Lond. B* **84**, 260–276 (1911).
2. Reguera, G. et al. *Nature* **435**, 1098–1101 (2005).
3. Kim, B.-H., Kim, H.-J., Hyun, M.-S. & Park, D.-H. *J. Microbiol. Biotechnol.* **9**, 127–131 (1999).
4. Cheng, S., Liu, H. & Logan, B. E. *Environ. Sci. Technol.* **40**, 2426–2432 (2006).
5. Reimers, C. E., Tender, L. M., Fertig, S. & Wang, W. *Environ. Sci. Technol.* **35**, 192–195 (2001).

AND WHAT IF IT WORKED?

In his satirical novel *Distraction*, science-fiction author Bruce Sterling imagines the development of a microbial power source for future cars — and the industry's fear of the far-reaching changes it heralds. Here, one of his characters rails against the hapless innovator:



You think it's easy running corporate R&D? It was just fine, as long as the guy didn't have anything. Jesus, nobody ever thought a goddamn sugar engine would work. The goddamn thing is just a germ in a box! We build cars up here, we don't build giant germs! Then they pull this crazy stunt and... well it makes our life impossible! We're a classic metal bending industry! We have interlocking directorates all throughout the structure, raw materials, fuel, spare parts, the dealerships... We can't get into the face of our fuel suppliers, telling them that we're replacing them with sugar water! We own our fuel suppliers! It'd be like sawing off our own foot!

"Batteries have the highest profit margin of

any automobile component. We were making money there. You can't make real money anywhere else in our business. The Koreans are building auto bodies out of straw and paper! We can't support an industry when cars are cheaper than grocery carts! What are we going to tell the unions? This is a great American tradition at stake here! The car defines America: the assembly line, suburbs, drive-ins, hot rods, teenage sex, everything that makes America great! We can't turn ourselves inside out because some big-brained creep has built an engine out of bug guts! There wouldn't be anything left of us! The guy is a menace to society! He had to be stopped.



BUSINESS

Japanese spin-offs face struggle for survival

Companies set up by academics are proliferating — but can they secure the investment they need to succeed? Ichiko Fuyuno investigates.

When Xueliang Song moved from northeast China to Japan 12 years ago, he was planning an orthodox research career at a university or a major corporation. But after Song earned his PhD in electronic engineering at the University of Tokyo in 2004, his lab head asked him whether he would like to set up a company based on his research. After mulling the idea over for a couple of days, Song decided it would be more fun than a conventional research career. “It sounded boring just to work at a big lab,” he says.

The company, Advanced Photonics, was established in March, and Song is already finding running a business and doing the optics research for product development a challenge. “Stepping into business from the academic world was like leaving for a foreign country again,” he says.

Song is part of a new wave of university researchers in Japan who have been encouraged to set up companies based on their ideas. As part of the government's efforts to pull the economy out of a decade-long slump, it has not just lifted restrictions on academics' business activities, but is offering them financial backing.

The change began in 2001, when the Ministry of Economy, Trade and Industry (METI) set the goal of doubling the number of university-based companies to 1,000 within three years. By March 2004, it said that number had hit 1,100 — today, the ministry says it is close to 1,300. This year, METI will spend ¥53 billion (US\$480 million) supporting the young companies through research, training and infrastructure grants.

But some outside observers are sceptical about whether many of these companies will ever prove to be viable — and even government officials concede that a lot of them are struggling to survive. Critics say that universities have also pushed companies into existence

without sound business plans, to help meet the government's target. And there is not much sign of private investment arriving to build the companies up.

A government report released in January revealed that half the university start-ups surveyed had no revenue and that only a dozen of them have gone public since 2002. Robert Kneller of the University of Tokyo, who is writing a book on venture capital in Japan and the United States, estimates that only about 50 biomedical start-ups in Japan, and fewer than 20 companies in other sectors, are likely to raise investments and achieve substantial sales.

Kneller and other critics say that it remains immensely tough for start-up companies to flourish in a country where little venture capital is available. The venture-capital industry in Japan says its total investment there last year was US\$1.8 billion, against US\$22 billion in the United States and US\$13 billion in Europe during 2004, the last year for which data are available. But some outside observers believe the true Japanese figure is much lower.

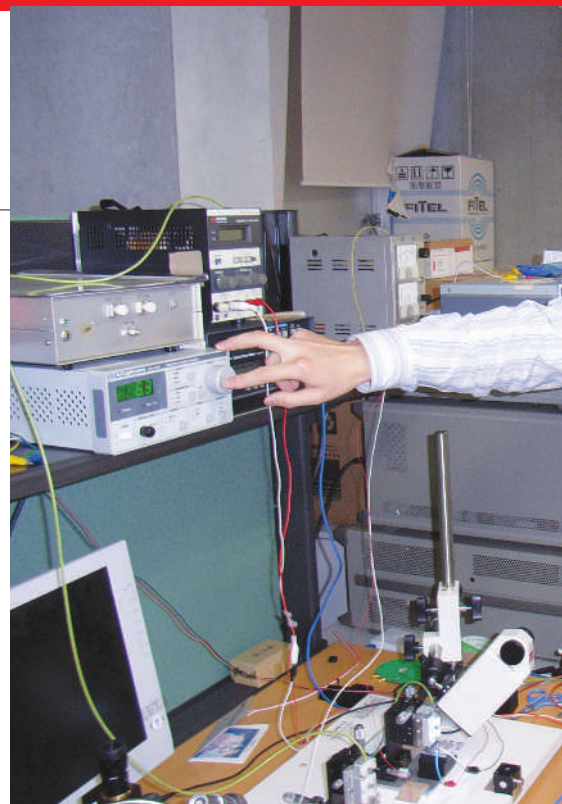
“The first stage of backing university start-ups is over. Now we'll only take care of competitive companies.”
— Masamichi Hashiguchi

Takafumi Yamamoto, chief executive of Toudai TLO, a technology-transfer unit at the University of Tokyo, says that venture-capital groups operating in Japan lack specialists who could help to nurture small companies. That's partly because their staff tend to

move regularly between different parts of the banks that own the firms. And the Mothers Index, the Tokyo equivalent of the Nasdaq stock exchange, only got going in 1999 and supports relatively few initial share offerings.

Such factors make the business climate difficult for start-ups. But the problems are exacerbated, economists say, because the most likely customers for new businesses — larger, existing businesses — prefer to stick with their established suppliers in Japan. The Japanese public sector doesn't have strong incentives to buy things from small companies, either.

With the number of university spin-offs



increasing, “the business environment is very severe”, says Hitoshi Kikumoto, a specialist in science and technology policy at the University of Tsukuba near Tokyo. In a survey, Kikumoto found that two-thirds of the start-ups had less than ¥30 million (US\$270,000) in working capital — most of which was provided by the researchers or their friends. This means that they won't be able to operate for long, he says.

“Few investors are willing to put money into a company at an early stage of development,” says Takuro Wakabayashi, a partner at the Advanced Science and Technology Enterprise Corporation (ASTEC), a Tokyo-based venture-capital group. ASTEC created a ¥1-billion investment fund in 2002, and has so far invested in 16 start-up companies — including Advanced Photonics. He believes that there are promising business ideas to be unearthed in Japan's universities, but that they need to be nurtured with great care. “Business opportunities are there,” he says. “But if we leave them alone, they'll quickly fail.”

In effect, researchers at the nascent companies have to be taught good business practices. Advanced Photonics is a case in point. The company aims to exploit Song's ideas for optical devices that will transmit data at very high speeds. Hiroaki Nakai, a former McKinsey management consultant who directs ASTEC, believes the device could compete with rival ideas being pursued by major corporations such as NEC. But on his frequent trips to Advanced Photonics' labs, Nakai has had to concentrate on teaching Song basic business skills such as how to compile a balance sheet. He now plans to hire a full-time chief executive for the company, leaving Song free to concentrate on his research.

Song, meanwhile, says he is learning more



L. FUYUNO

Xueliang Song (standing) is one of a number of academics now running spin-off firms in Japan.

business-like behaviour — such as keeping some of his ideas to himself, instead of sharing them with colleagues.

With backing from ASTEC, the company has now built a prototype device and hopes to commercialize it by 2010. Some of ASTEC's other companies are already farther down the commercial road. Perseus Proteomics, spun off from the University of Tokyo in 2001 to develop antibody-based drugs as possible cancer treatments, initially raised ¥600 million from ASTEC and other venture-capital groups. Wakabayashi effectively managed the company himself, until a full-time president was appointed in 2004. This February, Fujifilm bought a 22% stake in the company for ¥1 billion, and one of its antibodies is expected to enter clinical trials next year. Perseus has so far raised ¥2.4 billion in total.

The government has meanwhile taken several steps to help young companies. METI has marked out a dozen geographical 'clusters' nationwide and appointed a prominent academic or businessperson to supervise university spin-off businesses in each one. And the New Energy and Industrial Technology Development Organization (NEDO), a branch of METI, is providing scholarships for about 100 young would-be entrepreneurs at universities and research institutes.

But officials are aware of criticism that the support has been spread too thinly. "We can't support everyone," says Masamichi Hashiguchi, director of NEDO. "The first stage of backing university start-ups is over. Now we'll only take care of competitive companies."

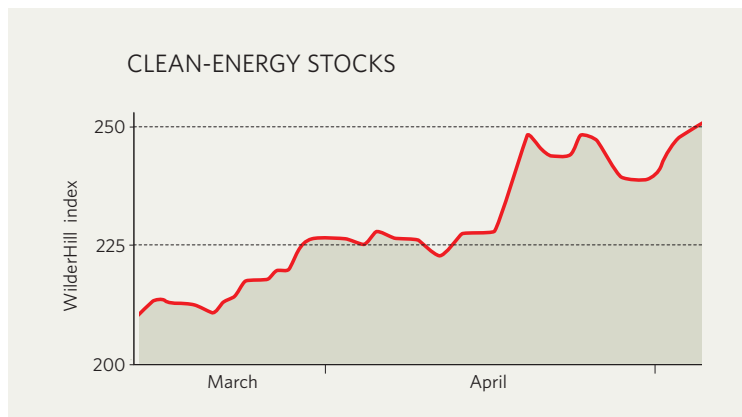
IN BRIEF

YEAST YIELD Merck is to pay \$400 million for a privately held biotechnology firm that improves drugs by fine-tuning their sugar structures. GlycoFi of Lebanon, New Hampshire, has developed technology that uses specially engineered yeasts to attach sugars to proteins with high accuracy. About 70% of drugs are proteins with sugars attached. GlycoFi was founded in 2000 and built up with \$35 million from venture-capital groups such as Polaris Venture Partners and SV Life Sciences. It will now become a wholly based subsidiary of New Jersey-based Merck.

NO CONCORDE Both Boeing and NASA have emphatically denied that they are planning talks with the Japanese Aerospace Exploration Agency (JAXA) on the joint development of a supersonic passenger plane. Officials in Japan had told newspapers there that they were hoping to meet with the Americans to discuss such a project, with a view to boosting innovation in their domestic aerospace industry. Japan already has a small, exploratory collaboration with France on supersonic airliners.

VAXGEN SETBACK Shares in California vaccine supplier VaxGen plunged by more than a third when it admitted it will be nearly two years late delivering 75 million doses of its recombinant anthrax vaccine to the US government. Under the terms of its \$878-million contract, which is part of Project BioShield, the company won't be paid until it delivers the vaccine. It now says that won't happen before late 2007 at the earliest, and it blames the delay on the government demanding more trials of the vaccine's safety and efficacy.

MARKET WATCH



A period of exuberance — irrational or otherwise — has sent the price of 'clean-energy' stocks shooting upwards by almost 50% since the beginning of the year.

Continuing worries about the high price of oil and a steady drumbeat of political and business endorsements for renewable energy have brought cash pouring into the clean-energy sector — and almost all stocks in it are rising.

"What we're seeing is that mainstream investors want clean-energy stocks," says Michael Liebrich of London-based New Energy Finance. "It's not exotic anymore."

Rob Wilder, whose company WilderShares in California keeps the WilderHill index (stock market symbol: ECO) of US-listed clean-energy stocks, says that the prevalent attitude towards the sector has reversed sharply. Not so long ago, he says, his was the "voice in

the wilderness" asking people to take clean energy seriously. "Now there's so much hyperbole, that I sometimes wish there were some naysayers. The more the sector is universally acclaimed, the more nervous I get."

Since a fund was launched last year that allows investors' money to track the WilderHill index, some US\$800 million has poured into it. But Wilder thinks some of the strongest growth prospects are in companies that are still too small to be represented in the index — including thermal (non-photovoltaic) solar energy, and ethanol production from cellulose.

Liebrich says he worries that market valuations of some stocks may have moved ahead of their real worth. But he argues that the overall rise in the market is "not a bubble" because the long-term prospects for the sector are strong.

Colin Macilwain

TGN1412: scrutinizing preclinical trials of antibody-based medicines

SIR — Your News story about the recent clinical-trial disaster, “Can super-antibody drugs be tamed?” (*Nature* **440**, 855–856; 2006), reports TeGenero founder Thomas Hünig as saying that the monkey CD28 receptor — the target of the TGN1412 antibody — is identical to the human one. TeGenero’s ‘investigator’s brochure’ (available via www.mhra.gov.uk) confirms that toxicity testing of TGN1412 was done in rhesus and cynomolgus monkeys.

Although TeGenero says that the amino-acid sequence of the critical portion of monkey CD28 (the C'D loop) is identical to that of human CD28, published amino-acid and DNA sequence data on rhesus CD28 indicate substantial genetic diversity within the species. Differences of up to 4% (9 of 220 amino acids) between rhesus and human proteins have been found (F. Villinger *et al.* *Immunogenetics* **53**, 315–328; 2001).

No data are available for CD28 in cynomolgus monkeys, but as they are closely related to rhesus monkeys, a similar degree of variation can be expected.

Two variable positions in the rhesus CD28 sequence — amino acid 65 (glutamic acid or glycine) and 104 (asparagine or tyrosine) — are located on the edge of the contact region between human CD28 and TGN1412. Although they may be tolerated, they are likely to substantially affect the strength of antibody–antigen interaction.

Identity of sequence in the C'D loop is also insufficient as a criterion for compatibility, as this segment contains only 6 of the 12 amino acids of CD28 that make contact with TGN1412. Substitutions at any position in this area might be expected to affect the shape of the antigenic site. Binding of the antibody to rhesus CD28 could therefore potentially be much weaker than its binding to human CD28. And differences in the binding affinities could alter the balance between activation of regulatory T cells (bearing relatively high levels of CD28), on the one hand, and T-helper cells and eosinophilic granulocytes (with lower CD28 expression), on the other (G. Woerly *et al.* *J. Exp. Med.* **190**, 487–495; 1999).

We propose that future preclinical studies of antibody-based drugs should be performed using a primate species in which the antigenic site matches that in humans completely, and that the individual primates used should be genotyped to exclude polymorphisms.

Furthermore, preclinical studies should also include comparative measurements of the binding affinities for both human

antigen and primate antigen, to control for unforeseen variations in protein structure.

Søren Hansen, R. Graham Q. Leslie

Medical Biotechnology Center and Institute of Medical Biology, University of Southern Denmark, Winslowparken 25, DK-5000, Denmark

Drug giants hamstrung by timid middle management

SIR — Your Business story “Too little, too late” (*Nature* **440**, 603; 2006) discusses a merger between Bayer and Schering, implying that the move is needed to make them large enough to succeed. My experience, during 50 years’ research in big pharma, is the opposite. Large companies are always inefficient because their command structure makes them so. Any organization with many layers, where power flows from the top down, works against innovation — look at the widely reported depletion of big-budget companies’ pipelines (see *Nature* **439**, 886–887; 2006).

The reason is that people in the middle layers, who neither control events nor engage in discovery, are too afraid to respond favourably to genuinely new ideas. If they encourage one and then it flops, as most innovations do, they are marked for demotion or dismissal. But if they kill novel programmes, nobody will ever know that a great thing died before it was born, and they are safe. The top layers cannot evaluate research themselves, because so much information is generated at the bench that they are forced to delegate most up-or-down decisions to middle management. Nowadays most of the innovation takes place in small outfits, because it is not crushed there.

Raymond A. Firestone

59 Barnes Road, Stamford, Connecticut 06902, USA

ID paper wasn't impartial enough for committee

SIR — As a former member of the Board of the Social Sciences and Humanities Research Council of Canada (SSHRC), I read with some irritation your News story “Doubts over evolution block funding by Canadian agency” (*Nature* **440**, 720–721; 2006).

The proposed study of the effects of popularization of intelligent design was, in fact, simply rejected and thus could not be “blocked” by the council. Given that the average success rate is less than 40%, the majority of applicants are bound to fail.

In this case, an excerpt from the rejection letter, lifted from its context, has been used to suggest that the committee felt there was inadequate “justification for the assumption

in the proposal that the theory of evolution, and not intelligent-design theory, was correct”. But this excerpt can be interpreted in a less dramatic manner: the committee simply thought the study was not impartial enough in its approach. After all, social-science research should study phenomena and not promote a particular view; many scholars legitimately demand a symmetric approach.

One can, of course, disagree with such comments, and many researchers do when they do not get their funding. But the author in this case is misguided in his view that the rejection “illustrates how the misunderstanding of evolution and intelligent design can go to all levels of Canadian society”.

The SSHRC does not interfere with peer-review committees. Instead of pointing his finger at the SSHRC, the researcher should complain about his peers who were not convinced by his proposal or, better still, he should learn from their comments.

Yves Gingras

Département d'histoire, Centre interuniversitaire de recherche sur la science et la technologie, Université du Québec à Montréal, case postale 8888, succursale Centre-ville, Montréal, Québec H3C 3P8, Canada

Supporting the use of artemisinin in combination

SIR — Your News story “Malaria breakthrough raises spectre of drug resistance” (*Nature* **440**, 852–853; 2006) raises concerns about a new source of artemisinin for treating malaria. The Institute for OneWorld Health (IOWH) strongly supports the World Health Organization guideline that uncomplicated *falciparum* malaria be treated with artemisinin combination therapies, and not by artemisinin alone as a monotherapy.

We are leading the product development effort for the Artemisinin Project in the developing world and will sell microbially derived artemisinin only for the manufacture of combination antimalarial therapies, except for treatments of severe malaria with parenteral formulations of artemisinin derivatives.

By leveraging the promise of synthetic biology, the IOWH and its partners hope to dramatically reduce the cost of antimalarials for the people who most need them.

Victoria G. Hale

Institute for OneWorld Health, 50 California Street, San Francisco, California 94111, USA

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

COMMENTARY

S. GROSSE/ELSCHKA/MSF



Improved rapid tests for malaria are needed to accurately diagnose infection and to avoid over-prescription of drugs.

Neglected tests for neglected patients

Alternative ways to develop diagnostic tools for use in resource-poor settings can, and do, exist, argue **Martine Usdin, Martine Guillermin and Pierre Chirac** of Doctors without Borders.

Organizations such as Doctors without Borders (Médecins Sans Frontières, or MSF) provide emergency medical assistance to populations that do not have access to medical care. In the various settings in which MSF works, the absence of suitable diagnostic tests is far more the rule than the exception. The need for reliable diagnostics is closely linked to the need for safe and effective treatments, and many of the diseases considered 'neglected' in terms of treatment are equally neglected in terms of diagnosis. In the case of sleeping sickness, for example, existing treatments are highly toxic and can kill up to 10% of patients¹; an accurate diagnosis must therefore be made before starting treatment. In the case of malaria — which kills more than 1 million people a year² — misdiagnosis of the disease from clinical signs alone^{3,4} is often a problem. The need to avoid overprescribing antimalarial drugs, to minimize costs and the risk of developing resistance, reinforces the need for an adequate diagnostic test.

MSF and other humanitarian organizations face similar challenges in diagnosing other major diseases in resource-poor settings,

notably tuberculosis, HIV in infants and HIV-associated infections such as cryptococcal meningitis. The lack of suitable diagnostic tests for use in the field means that the identification of many diseases is based solely on clinical symptoms. The long list of diseases with no field-adapted diagnostic tools includes leishmaniasis, shigella, typhoid and bacterial meningitis. MSF is attempting to address these needs by exploring how available technology can best be used, either by improving existing tests or by developing new, field-adapted ones.

What is needed?

The reasons for the gaps between need and availability are manifold. In some cases, existing tests are not sensitive or specific enough, or some of the fundamental biology of the disease is still not known. In other cases, technologies do exist but are too complex, too expensive or not suitable for remote settings with poor infrastructure. The technically challenging, painful and risky diagnosis of meningitis and sleeping sickness through lumbar puncture is a good example. The gaps between field needs and available solutions are large, but, as we

argue here, they can be effectively reduced by adopting a 'needs-based' approach to research and development and by making use of existing resources.

As with medicines for neglected diseases, the development of diagnostic tools for use in resource-poor settings is often thought to have little market incentive for private companies. Decision-making in the pharmaceutical and biomedical industries is increasingly driven by priority-setting based on expected market returns⁵. This priority-setting puts the needs of the company shareholders before those of the patient. The experience we document here shows how, by combining the specific needs identified by MSF in the field with existing scientific expertise, the development of appropriate diagnostic tools can be accelerated. By speeding up this process we can keep costs down, and thereby increase the likelihood that a test will be developed or marketed. Such opportunities should be attractive to both test providers and purchasers in the private and public sectors.

An example of such a needs-based product-development strategy is the development of a

new rapid diagnostic test (RDT) for malaria that is now in the final stages of field evaluation. Currently, MSF performs more than 4 million malaria RDTs a year, mainly in Africa and Asia. Specific problems with existing tests were previously identified by MSF staff. For example, the current 'gold standard', smear microscopy, requires trained technicians and careful quality-control, which is not always possible in the field. Existing RDTs based on detecting the malaria parasite protein HRP2 in fingerprick blood were seen to be useful and cost-effective, but are limited because HRP2 proteins circulate for months in the blood after cure and therefore the test cannot distinguish between an old, resolved infection and a new one. Nor do they detect all strains of malaria.

In 2003, MSF decided to coordinate the development of an alternative test, by starting with a precise definition of what was needed in the field — what type of test, who would use it, and how the information obtained would be integrated into subsequent medical decisions. It is tempting to make the most elegant, informative and cutting-edge device. However, the design goal was a test most adapted to the setting in which it was to be used — mainly in malaria-endemic areas in Africa, where heat and humidity, ease of use and cost are important concerns. The needs in Asia are different: there is more than one endemic parasite species and different levels of infection and resistance in the populations that make additional information, such as the strain of parasite, desirable.

Everyone's happy

MSF brought together scientists, field doctors and manufacturers who applied their expertise to the development of an alternative test based on the detection of the malaria parasite enzyme pLDH in the blood. This test is more suited for diagnosing malaria in Africa than the HRP2 test because, among other things, pLDH levels drop rapidly with clearance of parasites, allowing clinicians to distinguish between patients who are still infected with malaria after treatment, and those in whom fever symptoms have another cause. The possibility of a pLDH test was identified by MSF from published papers on the characterization of a panel of pLDH monoclonal antibodies⁶. After several rounds of discussions to define the medical needs of field clinicians and the integration of the results into patient care, MSF contacted the US developer of the antibodies, Michael Makler of the company Flow Inc. in Portland, Oregon, to explore the possibility of using his reagents.

The initial feasibility of the project was tested by other leading scientists who gave advice and performed *in vitro* validation. The antibodies were then offered to three manufacturers who worked to optimize the test format for ease of use and performance, and developed three independent versions of the pLDH test. These were then evaluated in MSF



Malaria tests tailored for use in Africa.

field projects. Opening up the market to more than one company encouraged the manufacturers to pursue development of a quality product that is adapted to field conditions, while maintaining the price at an accessible level. The entire development process, from identification of diagnostic needs through to field evaluation, was completed in less than two years, at the relatively low cost of about €80,000 (US\$100,000), most of which went into field trials and was provided by MSF. Development of a new test can typically cost many hundreds of thousands of dollars⁷.

This unusual development process has shown that products that are needed in the field, especially those for which there is not a 'traditional' or high-paying market, can still be developed in an efficient, commercially viable and mutually beneficial way. The development of this test differed from a typical diagnostic device in that low cost and field specifications were the main considerations from the outset, rather than market-driven (how much can we charge?) or technology-driven (what is the maximum information that we can deliver?) processes. The development team (MSF, Michael Makler, three different manufacturers, scientific and technical consultants, and field users) were guided at all times by these specifications. Thus, although all the players had their own motivations for participating, whether it was altruism, profit, need, curiosity or otherwise, all parties were working towards the same goal.

A successful model

The development of the malaria test built both on existing research (the use of pLDH as a marker for malaria infection) and existing technologies (immunochromatographic rapid tests). This created an efficient development model that we believe can be replicated elsewhere. In universities and industrial labs there are many — 80 °C freezers or bottom drawers of filing cabinets containing antibodies, constructs or designs that have been shelved owing to lack of funding, interest or conflicting corporate or research priorities. MSF and others can act as partners in guiding the development of these 'leftover' products into appropriate tools.

But large doses of goodwill and unlimited

access to freezers are not enough. In most cases, developing such tests requires cash and infrastructure, resources that are not always available. However, if the test is based on a solid assessment of the field needs and capabilities, it is more likely to be useful and the resources will be well spent. This happens when medical care drives the development of tests, rather than doctors being forced to adapt their medical care to the existing tools. The pLDH development model was a successful one, and it can and should be applied in other contexts.

MSF's experience during the development of this new malaria test shows how common interests can be fostered to bridge the gap between the needs of product developers and end users for the ultimate benefit of all involved. Given the long list of diseases with neglected diagnostics, there are many more opportunities to act and to make a difference. We appeal to scientists to respond to such opportunities by making existing reagents and ideas available for the development of diagnostic tests for these diseases. Such projects can make financial sense and have a huge impact on serious health issues, often without the enormous budgets of traditional development models.

We have described just one of many approaches needed to address the gaps in diagnostics; others include better use of existing reagents, training of local technicians in the field and a more creative and flexible approach to traditional product development.

We recognize that there are other factors that limit the development of new tools. In many cases, intellectual-property restrictions pose a significant barrier. There are also important issues in ensuring the quality of the tests and in developing international standards for diagnostic tools. However, we have shown that alternative models that are realistic about the needs and limitations in the field, and that take account of the different motivations of the people involved, do work. By linking existing resources with user needs, we can encourage the development of crucial diagnostic tools, for the benefit of developers and patients alike.

■
Martine Usdin, Martine Guillermin and Pierre Chirac are with the MSF Campaign for Access to Essential Medicines, 4 rue St Sabin, 75011 Paris, France.

1. Chappuis, F., Loutan, L., Simarro, P., Lejon, V. & Buscher, P. *Clin. Microbiol. Rev.* **18**, 133–146 (2005).
2. Carter, R. & Mendis, K. N. *Clin. Microbiol. Rev.* **15**, 564–594 (2002).
3. Luxemburger, C. et al. *Trans. R. Soc. Trop. Med. Hyg.* **92**, 45–49 (1998).
4. O'Dempsey, T. J. et al. *Trans. R. Soc. Trop. Med. Hyg.* **87**, 662–665 (1993).
5. *Project Selection, Decision Analysis and Profitability in the Pharmaceutical Industry* www.wma.net/e/publications/pdf/2001/senn.pdf
6. Piper, R. et al. *Am. J. Trop. Med. Hyg.* **60**, 109–118 (1999).
7. <http://grants.nih.gov/grants/guide/rfa-files/RFA-AI-95-001.html>

BOOKS & ARTS

From here to eternity

How do physicists cope with the concept of infinity in an expanding Universe?

The Infinite Cosmos: Questions from the Frontiers of Cosmology

by Joseph Silk

Oxford University Press: 2006. 256 pp.
£18.99, \$29.95

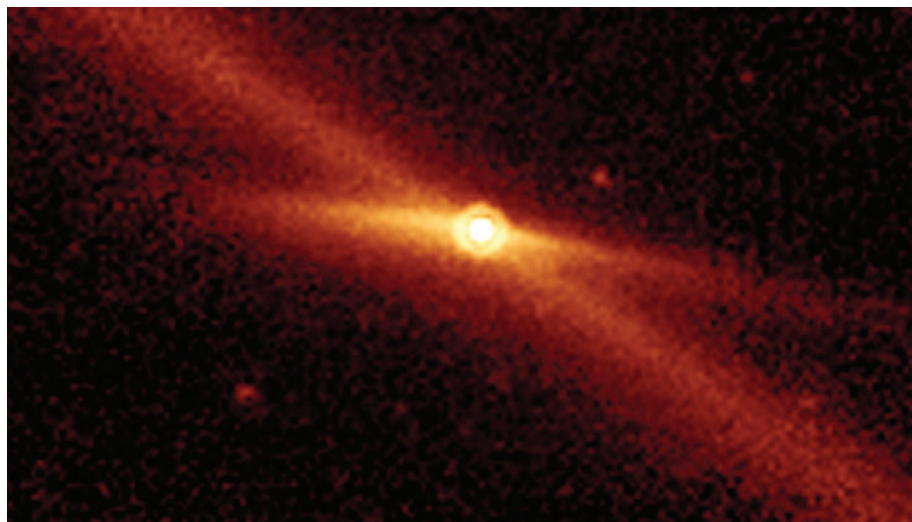
Peter Coles

Scientists usually have an uncomfortable time coping with the concept of infinity. Over the past century, physicists have had a particularly difficult relationship with the notion of boundlessness. In most cases this has been symptomatic of deficiencies in the theoretical foundations of the subject. Think of the 'ultraviolet catastrophe' of classical statistical mechanics, in which the electromagnetic radiation produced by a black body at a finite temperature is calculated to be infinitely intense at infinitely short wavelengths; this signalled the failure of classical statistical mechanics and ushered in the era of quantum mechanics about a hundred years ago. Quantum field theories have other forms of pathological behaviour, with mathematical components of the theory tending to run out of control to infinity unless they are healed using the technique of renormalization. The general theory of relativity predicts that singularities in which physical properties become infinite occur in the centre of black holes and in the Big Bang that kicked our Universe into existence. But even these are regarded as indications that we are missing a piece of the puzzle, rather than implying that somehow infinity is a part of nature itself.

The exception to this rule is the field of cosmology. Somehow it seems natural at least to consider the possibility that our cosmos might be infinite in extent or duration. If the Universe is defined as everything that exists, why should it necessarily be finite? Why should there be some underlying principle that restricts it to a size our human brains can cope with?

But even if cosmologists are prepared to ponder the reality of endlessness, and to describe it mathematically, they still have problems finding words to express these thoughts. Physics is fundamentally prosaic, but physicists have to resort to poetry when faced with the measureless grandeur of the heavens.

In *The Infinite Cosmos*, Joe Silk takes us on a whistle-stop tour of modern cosmology, focusing on what we have learned about the size and age of the Universe, how it might have begun, and how it may or may not end. This is a good



NASA/JPL-CALTECH/UNIV. MINN.

Very distant objects, such as the comet Encke, make it seem reasonable to think of an infinite Universe.

time to write this book, because these most basic questions may have been answered by a combination of measurements from satellites gathering the static buzz of microwaves left over from the Big Bang, from telescopes finding and monitoring the behaviour of immensely distant supernova explosions, and from painstaking surveys of galaxy positions yielding quantitative information about the fallout from the primordial fireball. Unless we are missing something of fundamental importance, these observations indicate that our expanding Universe is about 14 billion years old, contains copious quantities of dark matter in some unidentified form, and is expanding at an accelerating rate.

According to the standard model of cosmology that emerges, the Universe has a finite past and (perhaps) an infinite future. But is our observable Universe (our 'Hubble bubble') typical of all there is? Perhaps there is much more to the cosmos than will ever meet our eyes. Our local patch of space-time may have its origin in just one of an infinite and timeless collection of Big Bangs, so the inferences we draw from observations of our immediate neighbourhood may never tell us anything much about the whole thing, even if we correctly interpret all the data available to us.

What is exciting about this book is not that it is anchored by the ramifications of infinity, but that it packs so much into a decidedly finite space. Silk covers everything you might hope to find in a book by one of the world's

leading cosmologists, and much more besides. Black holes, galaxy formation, dark matter, time travel, string theory and the cosmic microwave background all get a mention.

The style is accessible and informative. The book also benefits from having a flexible structure, free from the restrictions of the traditional historical narrative. Instead there are 20 short chapters arranged in a way that brings out the universality of the underlying physical concepts without having too much of a textbook feel. The explanations are nicely illustrated and do not involve any mathematics, so the book is suitable for the non-specialist.

If I have any criticisms of this book at all, they are only slight ones. The conflation of the 'expanding Universe' concept with the Big Bang theory, as opposed to its old 'steady state' rival, is both surprising and confusing. The steady-state model also describes an expanding Universe, but one in which there is continuous creation of matter to maintain a constant density against the diluting effect of the expansion. In the Big Bang, there is only one creation event, so the density of the expanding Universe changes with time. I also found the chapter about God in cosmology to be rather trite, but then my heart always sinks when I find myself lured into theological territory in which I am ill-equipped to survive.

Peter Coles is in the School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

Lives after death

Alexander von Humboldt: A Metabiography

by Nicolaas A. Rupke

Peter Berg: 2005. 320 pp. \$38.95, £22.80, €32.50

Steven Shapin

It's said that you only live once, but a great scientist has many lives. These lives are not, however, personal existences but biographies, written by others, that make sense of the life and identify its historical significance. If the scientist is regarded as the founder of a discipline, or the author of a cherished theory or method, then biography can be a continuous creative act, changing in content and character as the science changes over time. Founding myths may be historically flawed but they often serve a function for the science concerned. And if the scientist is also a national hero, then even more is at play in these lives after death.

For Alexander von Humboldt (1769–1859), biographical life after death has been complex and contested. He was a polymath, or at least a poly-scientist; the only modern discipline that has plausibly laid claim to him as a founding figure is physical geography. His contributions to the field are marked by the names of the Humboldt Current along the west coast of South America, the Humboldt Glacier in Greenland, the Humboldt Mountain Range in Antarctica, and the Humboldt River and Humboldt Bay in the United States. But he also made contributions to natural history: a handful of plant and animal species are named after him, from the Humboldt penguin (*Spheniscus humboldti*) to the spectacular Humboldt lily (*Lilium humboldtii*). He was also a well-known traveller, diplomat, courtier, intellectual and popularizer of science.

Humboldt was a German who spent most of his productive life outside the country and wrote much of his scientific work in French. Together with his elder brother Wilhelm, a linguist who founded the University of Berlin, he has been a major German cultural icon, in the same pantheon as Goethe, Schiller and Beethoven.

But who was Alexander von Humboldt? Nicolaas Rupke doesn't seek a coherent answer to that question, nor does he think there can be one. He has not written a conventional biography but a life of lives, or 'metabiography', a meticulous study of how Humboldt's life was configured and reconfigured according to the sensibilities and needs of the changing cultural settings.

In the late nineteenth and early twentieth century, some German scientists saw Humboldt as a darwinian, even though he died before *On the Origin of Species* was published, and even though there is only patchy evidence of his dissent from the anti-evolutionist



A changing picture: perceptions of Alexander von Humboldt depend on the social context of the viewer.

views of such naturalists as Louis Agassiz.

In the years leading up to German unification under Bismarck, Humboldt was a hero to liberal nationalists, despite having served as royal chamberlain to a deeply conservative Prussian court. The Nazis did their best to make Humboldt into a philosophical idealist of their preferred mystical type and, as Rupke says, "a herald of the Third Reich's new socio-political order". The 'de-Nazifiers' of the postwar Federal Republic saw Humboldt as a model of global cosmopolitanism and tolerance, a "good German", and even an active friend of the Jews. In the German Democratic

Republic, Humboldt's early career as a mining engineer was used to embellish his proletarian credentials, and his remarks against slavery established him, according to an East German

text of the 1970s, as an opponent of "every form of racial discrimination, against every form of colonial expansion, and for the brotherhood of human beings of all colours".

More recently, the marxist Humboldt has been replaced with Humboldt the networking pioneer of globalized information economies, a 'green' Humboldt who viewed nature "as an ecological system", and a homosexual Humboldt whose long-suspected tendencies were 'outed' by several gay German writers, although this was disputed by scholars who either insisted on his heterosexuality or claimed that Humboldt had sacrificed sexual pleasures on the altar of scientific asceticism. Finally, the prisms of postmodernist scholarship have revealed a Humboldt who was a practitioner of disunified science and a man with no stable intellectual or political make-up at all, a ragbag of a man from whose life practically any evidence could be plucked to establish almost any identity, point or purpose.

Rupke is right to draw attention to the fact that shifting biographical traditions make one person have many lives, and his metabiography helps us to appreciate the historical instability of any scientific life, not just one as complex as Humboldt's. The past is never wholly a foreign country, and we make and remake past lives to suit our present purposes. No one whose life and work is of future interest can escape that fate. But one could also say that Rupke has given us a Humboldt just right for our own less certain and more self-conscious times — fractured, multiple and unstable.

Steven Shapin is Franklin L. Ford professor of the History of Science, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA.

A single survivor

Lonesome George: The Life and Loves of a Conservation Icon

by Henry Nicholls

Macmillan Science: 2006. 256 pp. £16.99, \$24.95

Rick Shine

Conservation biology is a peculiar hybrid discipline, bounded on one side by hard empirical science and on the other by ageing actresses cuddling cute baby seals. As a result, conservation priorities reflect a sometimes bizarre balancing act between knowledge- and emotion-based perspectives of the natural world.

This tension can occasionally result in an extraordinarily high public profile for an individual organism that can be argued to have some singular importance to conservation because it constitutes 'the last of its kind'. Some kind-hearted humans who are all too aware of the pain of loneliness on a personal level can readily empathize with the tragic circumstances of the last survivor. The animal can then take on iconic status — and (a cynic might say) attract publicity and financial resources out of all proportion to its actual conservation significance. The iconic individuals that evoke empathy from humans are

EXHIBITION

Martian arts

The outwardly dry, forbidding territory of science might seem infertile soil for the whimsy of the newspaper cartoonist. But judging by *Mars in their Eyes*, an engaging exhibition at London's newly opened Cartoon Museum, missions to our planetary neighbours are a notable exception. After all, everybody knows that martian soils are populated by little green men.

As the martian physiognomy is unknown, it is a gift to a cartoonist's imagination. The inhabitants of the red planet are also used to great effect as mostly perplexed, sometimes mischievous commentators on their Earthly neighbours' aspirations. In a 1976 cartoon from *The Philadelphia Inquirer*, for instance, a martian housewife bangs irritably on the ceiling of her underground condo as the Viking 1 probe drills through her roof.

Not all missions to Mars were as successful as Viking 1, however. The bad luck and cock-ups along the way are, inevitably, mercilessly lampooned. "Do we measure light years in centimetres or feet and inches?" wonders a white-coated, bespectacled scientist from the pen of David Haldane of *The Times* in response to one egregious example — the failure of NASA's Mars Climate Orbiter in 1999 following a disagreement between metric and imperial units.

The British-built Mars lander Beagle 2 (missing, presumed lost) features particularly prominently: its lead scientist, Colin Pillinger, is co-curator of the exhibition. A poignant item is



the pair of almost identical cartoons sketched in anticipation of Beagle's landing on Christmas Day 2003. Of these, only the one entitled *Silent Night* — not *Hark! Hear the Beagle Sing!* — ever went to press.

Whether as a potted history of Mars

exploration, a primer of the press's view of science, or simply as a bit of fun, *Mars in their Eyes* is well worth a visit. It can be seen at the Cartoon Museum in London, a block away from the British Museum, until 1 July.

R.W.

www.cartooncentre.com

BERNARD COOKSON

REUTERS



One foot in the grave: Lonesome George is thought to be the last of the Pinta Island tortoises.

usually mammals or birds, but one reptile has captured hearts and minds — the forlorn figure of Lonesome George.

George is an adult male Galapagos tortoise, and is believed to be the last surviving member of a race that was abundant on Pinta Island before being slaughtered as a protein source for the crews of passing vessels. Closely related organisms from different islands in this archipelago differ in size and shape, an observation that Charles Darwin famously used as evidence for his theories of evolutionary change. Indeed, it was an observation about the distinctiveness of tortoises from different islands that first alerted Darwin to the possibility of such spatial variation. So Lonesome George — captured in 1972, many decades after the Pinta race was thought to be extinct — is an icon for evolutionary biology as well as island conservation. As Henry Nicholls brings out in this delightfully written book, Lonesome George

has also served as a lightning rod for scores of political, economic and scientific debates that have raged around the Galapagos as the islands have been transformed from a quiet site of scientific focus to a mecca for international tourism.

The science surrounding George is murky. How should you treat an internationally known conservation icon? Maintain him in luxury, where the tourists can pay to see him and be inspired about the urgent need for conservation efforts? Or release him back to freedom on his island, with females of a closely related lineage, at the risk that he will fall down a crevice and die? Or try to extract some of those Pinta-specific genes so his lineage can live on — at the risk of causing harm to this taciturn and decidedly non-amorous tortoise?

George continues to plod through his luxurious enclosure over the decades as the battle rages around him, and people come and go.

These include a young Swiss biologist who massaged George's genitals on a daily basis for six months to try to arouse his sexual passions; a local fisherman who sends death threats against George whenever the government tries to regulate the harvest of marine fauna; scientists whose ideas and suggestions about managing George's affairs rarely manage to penetrate the regulations that surround such an icon; and a series of park directors determined to protect George at all costs.

Nicholls skilfully and seamlessly brings in a broad array of snippets from recent scientific research in a diverse array of disciplines as he outlines the potential approaches that a conservation biologist might take to George's plight. However, it is not even clear that George really is the last Pinta tortoise, as opposed to a recent reintroduction, because the genetic differences between the Pinta and Espanola tortoises are so small. Such local populations are blinking out all around us and rarely receive much attention from conservation biologists, let alone the public.

If George were a small brown lizard (or even worse, a snake), nobody would care. But the giant tortoises of the Galapagos convey a sense of serene contemplation and prehistoric dignity, and it is difficult not to anthropomorphize and imagine the world from George's viewpoint. The literary device of placing a reptilian icon at the centre of a dynamic play about science, conservation and our attitudes to nature results in a highly readable book that has much to say about the ways we flounder around in our attempts to protect things that seem important to us.

Rick Shine is at the School of Biological Sciences, University of Sydney, Sydney, New South Wales 2006, Australia.

Time for a change

Prokaryote: gene-sequence comparisons show the tree of life consists of bacteria, eukarya and archaea. The use of the term 'prokaryote' fails to recognize that an idea about life's origins has been proved wrong.

Norman R. Pace

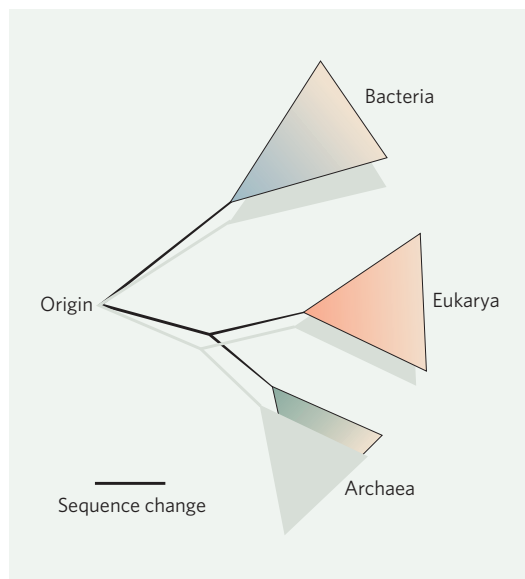
The explosive accumulation of gene sequences over the past few decades has brought a new perspective on life and its history. Some of the results indicate that we need to reassess our understanding of the course of evolution at the most fundamental level.

The current textbook paradigm for biological diversity and evolution is based on what I will call the prokaryote/eukaryote model. This posits that there are two kinds of cells: prokaryotic, those without nuclei (specifically, without nuclear membranes) and eukaryotic, those with a classical membrane-bounded nucleus. The model further posits that the former gave rise to the latter. The historical antecedents of this model are complex and rooted in the nineteenth century; for example, German biologist Ernst Haeckel positioned 'monera' (masses of protoplasm without a nucleus, later termed 'prokaryotes') at the base of his four-kingdoms phylogenetic tree.

The recognition that the main eukaryotic organelles, mitochondria and chloroplasts, were derived from bacteria by symbiosis between the bacteria and an ancestral eukaryotic cell prompted speculation on a similar origin for the eukaryotic nucleus. And the discovery of archaea — microbes that in many molecular ways resemble eukaryotes more than bacteria — resulted in proposals for archaeal origins for nuclear and cytoplasmic components of eukaryotic cells. Such proposals have sustained the concept that prokaryotes evolved into eukaryotes — an evolutionary model invoked by the terms themselves.

Molecular-sequence comparisons, first of ribosomal RNA genes in the late 1970s and of many other genes since, replaced analyses based on morphological subjectivities (such as the presence or absence of a nuclear membrane) with credible maps of evolutionary relationships between genes. These sequence comparisons have rendered the prokaryote/eukaryote model obsolete.

Ribosomal RNA, because of its ubiquity and slow rate of evolution, provides the most reliable view of the earliest evolutionary events. Comparisons of ribosomal RNA sequences show a three-domain tree of life (see figure). The diagram in essence is an experimentally derived map of biological organization and the course of evolution at the largest scale.



Comparisons of ribosomal RNA sequences reveal a three-domains tree of life, rendering the term 'prokaryote' obsolete.

Although some details of ribosomal RNA-based trees remain controversial, the basic three-domains structure and the relationships between the domains are generally accepted and are supported by observed biochemical variation. Phylogenetic trees based on all genes encoding the information-processing machinery needed to express genetic sequences are congruent with the three-domains tree. So the tree represents the evolutionary course of the genetic machinery, the functional core of genomes.

The lessons of the three-domains tree are profound. Instead of two kinds of organism, prokaryotes and eukaryotes, there are three: bacteria, eukarya (eukaryotes) and archaea. The root, or origin, of this universal tree, cannot be determined from ribosomal RNA sequences, but other phylogenetic results and biochemical correlates show that the genetic lines of eukarya and archaea have a common ancestral branch that is independent of that giving rise to the bacteria (see figure). That is, eukaryotes and archaea are more closely related to one another than either is to bacteria.

There is not a single (monophyletic) phylogenetic group upon which to hang the tag prokaryote. The major eukaryotic organelles, mitochondria and chloroplasts, are definitely bacterial in origin, but the nucleus is not. The nuclear line of descent is as ancient as the archaeal line

and not derived from either archaea or bacteria. Thus, the prokaryote/eukaryote model for biological diversity and evolution is invalid.

Some have asserted that 'prokaryote' has utility, as a term for non-eukaryotes. However, this is a negative and therefore scientifically invalid description; no one can define what is a prokaryote, only what it is not. Lumping bacteria and archaea conceptually discounts fundamental differences between those two kinds of organism and reinforces an incorrect understanding of biological organization and evolution.

But if we can't call them prokaryotes, what should we call them? That, of course, depends on what is meant by 'them'. If what is meant is 'the little stuff out there', then try microbes or microbial.

Other things may need more precision. For instance, references in textbooks and in the literature to 'prokaryotic transcription' — meaning a transfer of information from DNA to RNA that relies on sigma transcription factors — overlook a deeper complexity. Archaea do transcription differently from bacteria, with TATA-binding proteins, much like those used by eukaryotes. Bacterial transcription is the correct term to use in this case. Another such oxymoron is 'prokaryotic protein synthesis', when the bacterial version is meant.

I believe it is critical to shake loose from the prokaryote/eukaryote concept. It is outdated, a guesswork solution to an articulation of biological diversity and an incorrect model for the course of evolution. Because it has long been used by all texts of biology, it is hard to stop using the word, prokaryote. But the next time you are inclined to do so, think what you teach your students: a wrong idea.

Norman R. Pace is in the Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder 80309-0347, USA.

FURTHER READING

Woese, C. R. *Microbiol. Rev.* **58**, 1–9 (1994).
Woese, C. R. *Microbiol. Mol. Biol. Rev.* **68**, 173–186 (2004).
Sapp, J. *Microbiol. Mol. Biol. Rev.* **69**, 292–305 (2005).

NEWS & VIEWS

SOCIOBIOLOGY

The Phoenix effect

Kevin R. Foster

A spore-forming bacterium can escape from social collapse and extinction with a single mutation that has a dramatic effect. Here is evidence that a cooperative system can recover from the very brink of destruction.

Where there is society, there are cheaters that threaten to ruin it. The evolution of such selfish behaviour can destroy cooperation, and may even drive a species to extinction¹. When Fiegna, Velicer and colleagues (page 310)² mixed a cheater strain with a cooperative strain of a bacterium, therefore, tragedy seemed assured. Many populations of the bacteria did indeed die out but, in one, a new social strain arose phoenix-like from the social collapse. This new strain resisted the cheater, produced more spores than the original strain and, most amazingly of all, evolved these abilities through just a single mutational change.

The bacterium concerned is *Myxococcus xanthus*. It earns its title of 'social' bacterium from the ability of starving cells to aggregate and form a fruiting body of hardy spores, in which many of the cells die^{3–5} (Fig. 1). This social behaviour has been put to good use by Velicer and colleagues in an intriguing story that has spanned almost a decade^{3–5}. It began with an experiment that selected for mutant strains that grew rapidly under conditions in which they did not need to make spores⁵. Interestingly, several strains lost the ability to sporulate, and there was a further twist. Some could still form spores if mixed with the ancestral spore-former and, furthermore, produced more spores than the ancestor. Although socially inept when alone, these strains could exploit the shared resources in aggregations with the ancestor and use them to gain a selfish evolutionary advantage. Velicer and colleagues had inadvertently created cheater mutants⁴.

But would the cheaters completely replace the spore-former? To answer this, Fiegna and Velicer³ put mixtures containing one cheater strain and the social ancestor through several fruiting cycles in the lab by taking the spores from one cycle to seed the next. Two cheater strains persisted alongside the spore-former, but the more vigorous cheater rapidly increased in frequency and, as it did so, the results were devastating. By replacing the social ancestor, the cheater removed its own ability to produce spores. The result was population extinction: without spores, there was

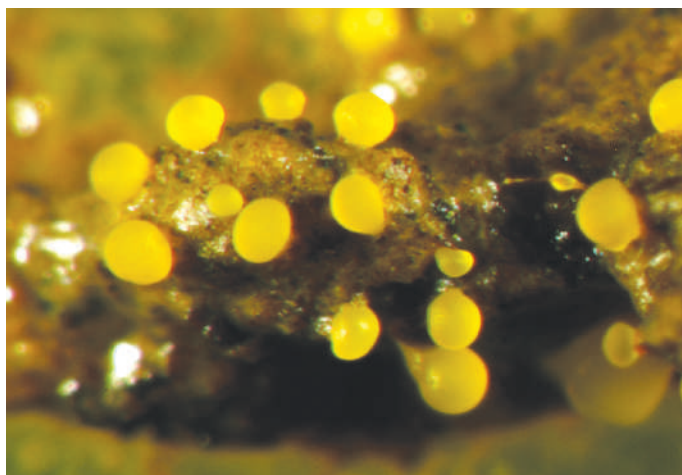


Figure 1 | Fruiting bodies of the social bacterium *Myxococcus xanthus* emerging from soil. Each fruiting body is about 100 μm in diameter, and contains a few thousand hardy spores that form through the aggregation of around 100,000 starving cells. (Photo by M. Vos.)

nothing to seed the next cycle of fruiting. These experiments provided a graphic illustration of what has come to be known as 'evolutionary suicide', where natural selection favours strategies that promote extinction. This process is alarming some conservation biologists, who worry that anthropogenic effects may be kick-starting spirals of selfish adaptations in some species that will drive their own destruction (Fig. 2b, overleaf)¹.

At least for the bacteria, however, all was not doom and gloom. Curiously, in one experiment, spores reappeared and the population recovered². It turned out that a new super-strain had evolved that could resist the cheater. This was named Phoenix after the mythical burning bird that can rise from its own ashes. In a tour de force of genetics, which involved the marathon task of sequencing the entire genome of the new strain, Fiegna *et al.*^{2,6} found that Phoenix arose with just a single mutation. This mutation simply increases the levels of one particular enzyme (an acetyl-transferase), but this is hypothesized to trigger a flood of subsequent changes in gene regulation and to drive a previously unrecognized route to sociality in *M. xanthus*.

To show that a social adaptation can escape from such a severe cheater, and can do so through a single mutation, makes this a landmark study^{2,6}. As with all studies, however,

there are caveats. Most importantly, this was all done in the lab, and strains were selected and mixed in ways that are unlikely to occur in nature. For example, natural aggregations of *M. xanthus* may only ever contain a single strain, which would mean that the lab cheaters that cannot develop alone would never spread. One must be cautious, therefore, when using these results to draw conclusions about the natural world. Nevertheless, the study represents a notable proof-of-principle that has intriguing implications for sociobiology.

First of all, it is quite amazing to find that a single mutation can drive such a dramatic social recovery, a result that underlines just how little we know about the genetic basis of social traits⁷. It is true that during a rapid population crash there may be little time for anything other than the most simple of genetic changes. However, a recovery mutation might also have resulted in only a slight initial improvement in spore production and a rather different evolutionary prognosis (Fig. 2c). If a new cooperator evolved to replace the cheater that was a poor sporulator, the way might be open for the original cooperator to re-evolve. This would mean the cheater could then re-invade, and perpetuate a potentially endless evolutionary game of rock-paper-scissors. Although this may seem to stretch credulity, such outcomes are known from both chemical

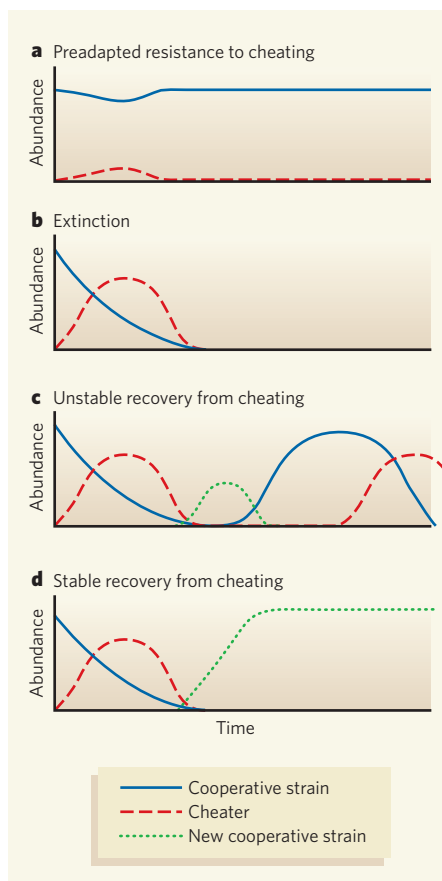


Figure 2 | Four possible outcomes when a cheater evolves in a social species. A cheater is an organism that exploits a cooperative adaptation for selfish gain. **a**, Preadapted resistance to cheating. It is typical to assume that social systems arise in such a way that cheaters can have only a limited impact (as shown), or do not succeed at all. Examples of preadaptations include high relatedness and pre-existing constraints that link cheating to a cost to the cheater¹¹. Policing and enforcement systems may evolve later to further constrain cheaters¹². **b**, Extinction. The cheater causes extinction of the social trait, or species (evolutionary suicide¹⁰). This selects for species preadapted to resist cheating¹⁰. **c**, Unstable recovery. A social strategy arises that resists the cheater but cannot out-compete the original strategy. The original strategy may reinvade and perpetuate a cycle of reinvasions in a rock–paper–scissors dynamic^{8,9}. **d**, Stable recovery. Sociality is restored by a strategy that out-competes both the cheater and the original strategy, as occurred with the Phoenix mutant². The result is a stable adaptation that protects the social system from the cheater. This process may be behind the policing and enforcement systems in other social species¹².

warfare in bacteria⁸ and male–male competition in lizards⁹.

This is not the case for Phoenix (Fig. 2d), which can out-compete the original strain and forms spores just fine. Such precipitous recoveries may turn out to be part of the evolutionary process. Extinctions play a key role in the history of life by removing species that are poorly adapted to persist, a process that favours both sexual reproduction and reduced

within-species conflict^{10,11}. However, the study by Fiegna *et al.*² shows that species may also escape from the very brink of disaster. There is some sense in this: a virulent cheater that threatens a population will necessarily result in strong natural selection for strategies that can re-evolve sociality in its wake. It was this effect that led to the rapid dominance of the Phoenix mutant, and one can speculate that it might also explain adaptations that police cheating in other societies¹². The final twist in the tale is that Phoenix actually produces more spores than the original strain. Escape from a virulent cheater is not just possible; it can even improve things.

Kevin R. Foster, currently in the Program for Evolutionary Dynamics, Harvard University, will shortly move to the Bauer Center for Genomics Research, Harvard University,

Cambridge, Massachusetts 02138, USA.

e-mail: kfoster@fas.harvard.edu

- Rankin, D. J. & López-Sepulcre, A. *Oikos* **111**, 616–619 (2005).
- Fiegna, F., Yu, Y.-T. N., Kadam, S. V. & Velicer, G. J. *Nature* **441**, 310–314 (2006).
- Fiegna, F. & Velicer, G. J. *Proc. R. Soc. Lond. B* **270**, 1527–1534 (2003).
- Velicer, G. J., Kroos, L. & Lenski, R. E. *Nature* **404**, 598–601 (2000).
- Velicer, G. J., Kroos, L. & Lenski, R. E. *Proc. Natl Acad. Sci. USA* **95**, 12376–12380 (1998).
- Velicer, G. J. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 8107–8112 (2006).
- Robinson, G. E., Grozinger, C. M. & Whitfield, C. W. *Nature Rev. Genet.* **6**, 257–270 (2005).
- Kirkup, B. C. & Riley, M. A. *Nature* **428**, 412 (2004).
- Sinervo, B. & Lively, C. M. *Nature* **380**, 240 (1996).
- Nunney, L. in *Levels of Selection in Evolution* (ed. Keller, L.) 238–252 (Princeton Univ. Press, 1999).
- Foster, K. R., Shaulsky, G., Strassmann, J. E., Queller, D. C. & Thompson, C. R. *Nature* **431**, 693–696 (2004).
- Frank, S. A. *Evolution* **57**, 693–705 (2003).

EXTRASOLAR PLANETS

A neptunian triplet

David Charbonneau

Three planets of Neptune mass have been discovered orbiting a Sun-like star known to have an asteroid belt. Exquisite measurements suggest that the search for habitable planets might be easier than assumed.

Our thirst for knowledge of planets orbiting stars similar to the Sun is tempered by the technological challenges of detecting them. We cannot see analogues of the Solar System directly; rather, the presence of extrasolar planets is inferred through effects that they induce on their parent star. The Doppler method, whereby astronomers search for subtle, periodic changes in the apparent speed of a star that result from its gravitational dance with an unseen planetary companion, has yielded all but a handful of the more than 180 known extrasolar planets¹. Heavier planets produce larger stellar wobbles, so it is not surprising that most of the worlds discovered so far have more in common with the distant gas and ice giants of the Solar System (Jupiter, Saturn, Uranus and Neptune) than with the smaller, closer terrestrial planets from Mercury to Mars.

But as techniques have been refined, so planets of lower mass have been revealed in increasing numbers¹. The current bestiary of extrasolar planets is therefore far from comprehensive. On page 305 of this issue, Lovis and colleagues² report unprecedentedly precise observations of the nearby, Sun-like star HD 69830. The fruit of their efforts is not one, but three orbiting planets (Fig. 1). The discovery is exciting for two reasons. First, the authors' technological advance implies that further low-mass planets will be spotted orbiting other stars. Second, the architecture of this particular planetary system bears some

intriguing similarities to that of our own Solar System.

The newly found planetary system is remarkable in that it possesses three planets located on nearly circular orbits within 1 astronomical unit of the star (1 AU is the Earth–Sun distance). The same is true of the Solar System. Where the HD 69830 system differs, however, is that the masses of the worlds range from 10 to 18 times that of Earth, and so are similar to that of Neptune. In the Solar System, the division between the low-mass terrestrial planets and massive gas giants was determined by the 'ice-line'. This is the distance beyond which the temperature in the protoplanetary nebula — the reservoir of gas and dust from which the planets formed — dipped below the freezing point of various hydrogen compounds. Beyond this point, much greater amounts of solid material, and so planets of much greater mass, were created.

The formation history of the HD 69830 system is thus a puzzle deserving of detailed study. Lovis and colleagues present² a preliminary calculation to show that the inner planets probably formed inside the ice line, and thus are likely to be predominantly rock, not gas. Their large masses require that HD 69830's protoplanetary nebula contained a larger quantity of solid material than did that of the Solar System. That inference is at odds with the observation that the star itself actually has a lower abundance of heavier elements, the stuff of planets.

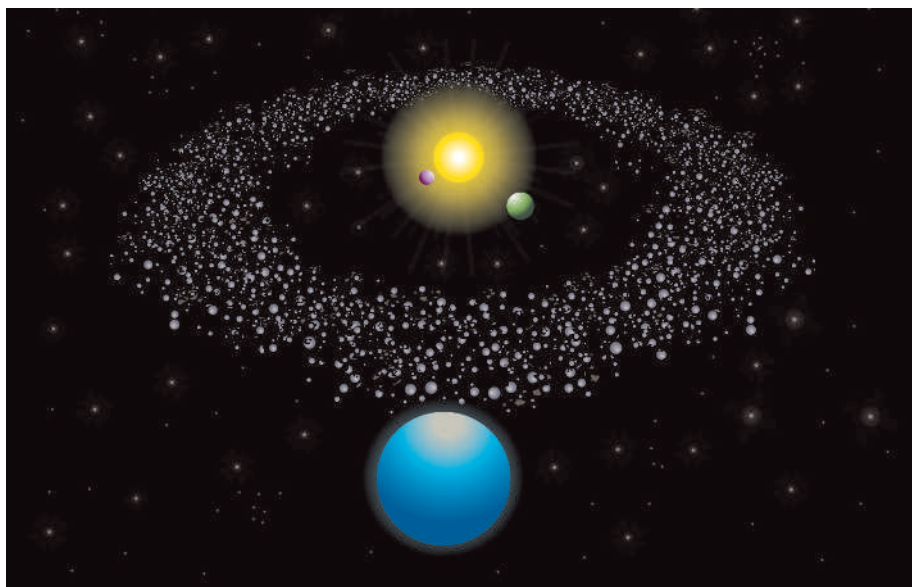


Figure 1 | Three's company. How the planetary system of the Sun-like star HD 69830 might look, according to Lovis and colleagues². Three Neptune-mass planets orbit the star on near-circular orbits at around 0.08 AU, 0.19 AU and 0.63 AU (where 1 AU is the distance from Earth to the Sun). Considerations of the gravitational influence of the three planets puts the most likely position for an asteroid belt, the existence of which has been inferred by measurements of infrared radiation³, at between 0.3 and 0.5 AU.

HD 69830 is no stranger to the spotlight. Last year, researchers using the NASA Spitzer Space Telescope announced that it probably possesses an asteroid belt³ — the only star similar in mass and age to the Sun that is known to have one. This conclusion came from the observation that the star was brighter than expected at infrared wavelengths, suggesting the presence of grains of dust that were small and were located within 1 AU of the star. At that distance, small grains could not survive very long, as the star's light would cause them either to fall inwards towards the star, or be blown outwards. The researchers therefore inferred that the grains must be constantly replenished by material spun off in collisions of large bodies in an asteroid belt. Notably, the implied mass of HD 69830's asteroid belt is roughly 25 times that of our own, seemingly in line with the beefed-up values of the planetary masses.

Intriguingly, these researchers also posited³ the existence of an unseen planet that quietly shepherded the asteroid belt in its orbit. It would seem that Lovis *et al.* have found the (plural) shepherds. Working with the converse logic, they consider² the gravitational influence of their three planets on an asteroid belt, and find that its position must be constrained to be either close to the star (0.3–0.5 AU) or far from it (beyond 0.8 AU). The earlier infrared observations favour the former location, which places the belt between the orbits of the centre and the outermost planet, but whether this location is truly stable and consistent with the infrared observations remains an open question. In the Solar System, the asteroid belt lies near 2.6 AU, between the orbits of Mars and Jupiter. The difference in its location in the HD 69830 system is surely a clue to differences in that star's planet-formation history.

One of the great quests of astronomy is to discover a small, rocky, Earth-like planet orbiting within the 'habitable zone' — the range of distances from a star for which the ambient temperature would permit a planet with liquid water and, perhaps, life as we know it. For the Sun, which is comparatively large and hot for our Galaxy, this orbital distance is great enough that the Doppler wobble induced by Earth would be a measly 9 cm s^{-1} , nearly an order of magnitude below even the exquisite measurement precision established by Lovis and colleagues. But most neighbouring stars are significantly less massive and cooler than the Sun, so a search for habitable planets using the Doppler method is feasible for two reasons. First, the lower stellar mass means an Earth-mass planet will cause a larger

Doppler wobble. Second, the reduced stellar brightness shifts the habitable zone inwards, reducing the orbital period of a habitable planet and further increasing the Doppler wobble. Should astronomers succeed in transferring the precision achieved by Lovis *et al.*² to low-mass stars, they might just turn up a few planets akin to our own.

What the Doppler method will, unfortunately, never reveal is the composition of the planets it detects. An exciting synergy therefore surely exists between such high-precision Doppler measurements and measurements to be made by two planned satellite missions, the COROT mission⁴ led by the French space agency CNES, and NASA's Kepler mission⁵. Both spacecraft will seek to identify rocky planets by the complementary method of transits⁶, in which a planet is revealed as it crosses the face of its parent star by a minute dimming of that star's light. By observing both Doppler signals and transits, both the mass and physical size of a planet can be estimated. That in turn yields a density and (by inference) a composition. Previous limits on the Doppler method implied that any terrestrial planets detected by COROT and Kepler would not receive corroborative Doppler measurements. Excitingly, Lovis and colleagues² have given us cause to rethink this pessimistic assumption. ■

David Charbonneau is in the Department of Astronomy, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA. e-mail: dcharbonneau@cfa.harvard.edu

1. Udry, S., Fischer, D. A. & Queloz, D. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) (Univ. Arizona Press, Tucson, in the press).
2. Lovis, C. *et al.* *Nature* **441**, 305–309 (2006).
3. Beichman, C. A. *et al.* *Astrophys. J.* **626**, 1061–1069 (2005).
4. Bordé, P., Rouan, D. & Léger, A. *Astron. Astrophys.* **405**, 1137–1144 (2003).
5. Basri, G., Borucki, W. J. & Koch, D. *New Astron. Rev.* **49**, 478–485 (2005).
6. Charbonneau, D., Brown, T. M., Burrows, A. & Laughlin, G. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) (Univ. Arizona Press, Tucson, in the press).

MICROBIOLOGY

Antibiotic stops 'ping-pong' match

Eric D. Brown

As bacteria become resistant to existing drugs, there is a need for antibiotics with new modes of action. Such a compound has been found, and it works by binding to an intermediate in the catalytic cycle of its target.

Pathogenic bacteria have developed strains that are resistant to almost all antibiotics in use today. Particularly worrisome are infections by a large group of bacteria classified as being Gram-positive, such as staphylococci and enterococci, which cause pneumonia and other, often fatal, infections. The problem is highlighted by the emergence of multiply-drug-

resistant strains of these organisms — so-called superbugs — that are resistant to vancomycin¹, a drug widely recognized as the last line of defence in many Gram-positive bacterial infections. On page 358 of this issue², Wang and colleagues report the discovery of a new antibiotic, platensimycin, that has potent antibacterial activity against these Gram-positive pathogens.

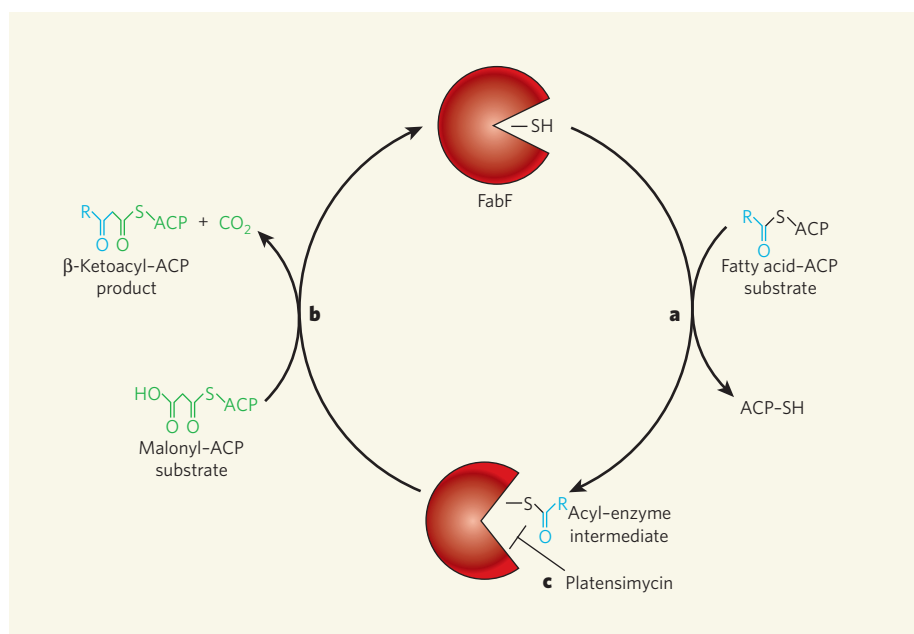


Figure 1 | The catalytic cycle of FabF for fatty-acid biosynthesis in bacteria. **a**, The fatty-acid fragment (in blue; R represents a general hydrocarbon chain) is transferred from a sulphur atom on acyl carrier protein (ACP) to a sulphur atom on cysteine in the FabF active site. In response, the active site becomes more open. **b**, Malonyl-ACP substrate (in green) binds to the acyl-enzyme intermediate, and loses carbon dioxide. The fatty-acid fragment is transferred from the active site, adding to the remains of the malonyl-ACP substrate, and thus extending the chain. **c**, Wang *et al.*² show that platensimycin inhibits the cycle by binding to the acyl-enzyme intermediate, preventing the addition of malonyl-ACP substrate — blocking fatty-acid biosynthesis, and so acting as an antibiotic.

In the past 40 years, only two antibiotics representing new chemical classes have reached the clinic, namely linezolid (an oxazolidinone³) and daptomycin (a lipopeptide⁴). Most classes of antibiotic were discovered in the 1940s and 1950s, and are directed at a few specific aspects of bacterial physiology — mainly biosynthesis of the cell wall, and of DNA and proteins. Subsequent tweaking of these chemical scaffolds has produced most of today's antibiotics. It is believed that widespread drug resistance among bacterial pathogens is due to the limited choice of antibiotics that exploit a relatively narrow range of mechanisms. Wang and colleagues² report² of a compound representing a novel class of antibiotic with activity against Gram-positive bacterial pathogens is thus particularly exciting, with the added bonus that platensimycin is effective against multiply-drug-resistant strains of staphylococci and enterococci.

The report reads like a textbook of modern antibacterial drug discovery, beginning with a screen of 250,000 extracts from drug-producing microorganisms. What follows is a series of elegant studies, spanning bacterial genetics, biochemistry, pharmacology and structural biology, and leading to the discovery of a small molecule derived from *Streptomyces platensis* that targets a seldom-exploited weakness in bacteria: fatty-acid biosynthesis. Fatty acids — organic acids with long hydrocarbon chains of between 8 and 18 carbon atoms — are the building-blocks of cell membranes and

bacterial surfaces. They are produced through the repetitive action of biosynthetic machinery that elongates the chains two carbon atoms at a time. The target of platensimycin is a key enzyme component of this machinery: β -ketoacyl-ACP (acyl carrier protein) synthase, also known as FabF.

The FabF enzyme mediates a reaction that has a fascinating catalytic cycle (Fig. 1). The fatty acid to be elongated is first transferred from ACP to FabF, leaving the fatty acid bound to the enzyme through an active-site amino acid (cysteine), affording a transient acyl-enzyme intermediate⁵. The source of extender carbon atoms is a molecular fragment known as a malonyl group, which is attached to ACP. The malonyl-ACP substrate is bound by FabF, and then loses carbon dioxide to produce a reactive two-carbon unit. The reactive unit attacks the acyl-enzyme intermediate and yields an elongated product that is released from the enzyme. Enzymologists describe this series of reactions as 'ping-pong'. This loose analogy refers to the way that the first substrate 'pings' into the active site and the first product 'pings' out, leaving the enzyme altered, so that the second substrate does the same as the first, leading to a product known as β -ketoacyl-ACP (Fig 1). There is thus a strict order involving the addition of fatty acid-ACP and then malonyl-ACP.

Wang *et al.*² provide an account of some terrific detective work that revealed that platensimycin binds only to the acyl-enzyme intermediate. Such intermediates are short-lived,

typically with lifetimes of the order of milliseconds, and so are difficult to observe. To overcome this problem, the researchers created a mimic of the acyl-enzyme intermediate. They prepared a variant of the FabF enzyme in which the cysteine of the active site was replaced with the amino acid glutamine. The chemical group that forms the side chain of glutamine mimics a bound fatty acid. The variant enzyme bound platensimycin with high affinity, and a high-resolution crystal structure of the complex of variant FabF with the new antibiotic was obtained. The structure reveals that formation of the acyl-enzyme intermediate is accompanied by structural changes that open up the active site, permitting binding of platensimycin in such a way that it blocks the addition of malonyl-ACP.

Platensimycin is a significant new antibacterial compound with an extraordinary mechanism, but it is not the first antibiotic known to inhibit bacterial fatty-acid biosynthesis^{6,7}. Isoniazid and triclosan are synthetic compounds that also target this pathway. Isoniazid has a clinical niche in treating tuberculosis in combination with other antibiotics, whereas triclosan has been widely used in soaps and plastics. In addition, cerulenin and thiolutomycin are natural products derived from fungi that target the same specific biosynthetic reaction as platensimycin in bacteria. That neither of these fungus-derived compounds has found a use in the clinic is testimony to the high standards required for a successful new antibiotic. Pharmaceutical companies have generally retreated from the field of antibacterial drugs, concentrating instead on chronic diseases with perceived product development and market advantages⁸. It is heartening, therefore, that platensimycin has been discovered and characterized by workers at Merck.

Wang *et al.* show that this antibiotic is effective in a mouse model of infection. The path ahead remains a long one that includes further preclinical study, and, if these studies are successful, extensive clinical trials for safety and efficacy in humans. Platensimycin is nevertheless the most potent inhibitor reported so far for FabF, and thus its discovery is an encouraging one. ■

Eric D. Brown is in the Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada.
e-mail: ebrown@mcmaster.ca

1. Weigel, L. M. *et al. Science* **302**, 1569–1571 (2003).
2. Wang, J. *et al. Nature* **441**, 358–361 (2006).
3. Ford, C. W., Zurenko, G. E. & Barbachyn, M. R. *Curr. Drug Targets Infect. Disord.* **1**, 181–199 (2001).
4. LaPlante, K. L. & Rybak, M. J. *Expert Opin. Pharmacother.* **5**, 2321–2331 (2004).
5. Price, A. C., Rock, C. O. & White, S. W. *J. Bacteriol.* **185**, 4136–4143 (2003).
6. Campbell, J. W. & Cronan, J. E. Jr *Annu. Rev. Microbiol.* **55**, 305–332 (2001).
7. Heath, R. J., White, S. W. & Rock, C. O. *Appl. Microbiol. Biotechnol.* **58**, 695–703 (2002).
8. Projan, S. J. *Curr. Opin. Microbiol.* **6**, 427–430 (2003).

SOLID-STATE PHYSICS

When is a metal not a metal?

Steven C. Erwin

When it's an insulator, of course. Materials that should in theory conduct electricity — but don't — are well known, but the anomalous behaviour of one material has caused particular head-scratching.

Every student knows the difference between a metal and an insulator: one conducts electricity and the other doesn't. Things get more interesting if you ask how this difference arises. Although the question is disarmingly simple, a rigorous answer was not available until about 1930, when Felix Bloch and Alan Wilson used the new quantum mechanics to create a theory that distinguished metals from insulators^{1,2}. The spectacular success of this 'band theory of solids', as it is now known, has made it a cornerstone of the modern theory of solids.

In a few glaring cases, however, band theory doesn't get it right: the materials known as Mott insulators, for instance, are experimentally insulating, yet band theory firmly insists they are metallic. Writing in *Physical Review Letters*, Cortés and colleagues³ take a fresh look at one such material — a germanium surface with a layer of adsorbed tin atoms — that has been puzzling theorists and experimentalists alike. Despite expectations based on analogy with similar materials, previous experiments on this system had failed to establish it as a Mott insulator. The results of Cortés *et al.* finally put this piece of the puzzle into place.

Exceptions to band theory provide fertile soil for testing new ideas about the behaviour of electrons in solids. Perhaps no one contributed more than Nevill Mott, a 1977 Nobel laureate in physics. Mott pointed out that the central approximation of band theory — that each electron moves independently, feeling the effects of the others only on the average — would fail badly in certain circumstances⁴. Armed with this realization, physicists could finally begin to understand those materials that owe their insulating nature to correlations in the motions of different electrons. These correlations arise from nothing more than the classical Coulomb repulsion between particles of like charge. But they can be decisive in materials in which the two competing tendencies of electrons are already in delicate balance: the desire to be spatially localized to minimize Coulomb repulsion, and the need to be delocalized to minimize the cost in kinetic energy from spatial confinement.

In bulk materials, the tendency to delocalize — and so conduct — usually prevails, because the freedom offered by three dimensions is greater than the Coulomb penalty. Not so at the surfaces of semiconductors, where electrons are effectively confined to two

dimensions (Fig. 1). Indeed, Mott insulators have been created over the past decade on the surfaces of many common semiconductors, including silicon carbide, gallium arsenide and even silicon itself. In most cases, adsorbed dopant atoms were used to tune the number of electrons to the point where band theory would predict a metal. When the evidence showed otherwise, Mott's electron correlations were a natural suspect.

Despite the occasional failure, Bloch–Wilson band theory still provides a useful language for understanding even those materials it gets wrong. Electrons in crystalline solids cannot have arbitrary energies, nor can they all have the same energy. Instead, the specific arrangement of atoms within the solid dictates the ranges of allowed energies — the

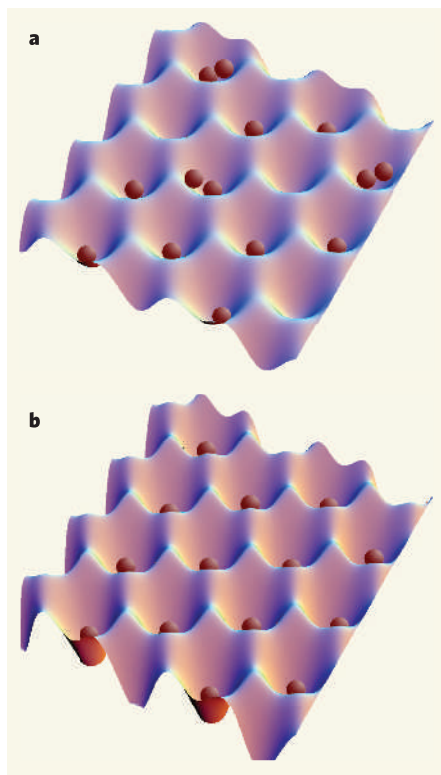


Figure 1 | Metal-insulator transition. **a**, The regular potential wells of a normal metallic state of a material with, on average, one electron per atom. **b**, If the confining potential is a little stronger (deeper wells), electrons find it harder to delocalize and so do not conduct — despite what the band theory of solids predicts. The material is a Mott insulator, such as the tin-covered germanium investigated by Cortés and colleagues³.

'bands' of band theory. The electrons must occupy the available bands sequentially, starting with the energetically lowest and proceeding upwards. Each band can accommodate two electrons of opposite spin. It follows that if the solid (more properly, one unit cell of the solid) contains an odd number of electrons, then at least one band must be incompletely filled. Such a material is guaranteed to be metallic within band theory.

The experimental search for surfaces meeting these conditions, but nonetheless exhibiting insulating behaviour, began in earnest in the 1990s. In one classic experiment, a Mott insulator was engineered on a silicon surface by first depleting its electrons using implanted boron. Then potassium was adsorbed until just enough electrons were supplied to meet the Bloch–Wilson criterion for a metal⁵. But data from photoemission spectroscopy (which uses light to eject electrons from a material and reveal their energy distribution) showed a gap between filled and empty bands. This contradiction with band theory provided strong evidence, eventually confirmed theoretically⁶, that the potassium-covered silicon surface is a Mott insulator.

A similar strategy was also applied to the surface of germanium, but with important differences. Instead of the boron–potassium combination, a single overlayer of lead was used to supply the same number of electrons. When this lead–germanium system was cooled below 100 kelvin, a small energy gap appeared, but accompanied by a new feature: a periodic, rumpling distortion of the lead overlayer⁷. When tin atoms (with the same number of valence electrons as the lead atoms) were added instead, the same rumpling was found, but the gap was mysteriously absent⁸. It has been persuasively argued⁹ that the rumpling seen at low temperatures in both systems is really the freezing of a vibrational mode in which, at higher temperatures, the tin atoms undergo rapid up–down oscillations. But the larger question — why the tin–germanium system seemed not to be a Mott insulator — remained unresolved.

Cortés and colleagues' contribution³ to the story is twofold. First, they establish that tin-covered germanium is indeed a Mott insulator, but only at very low temperatures. Second, they present angle-resolved photoemission data that give an especially detailed picture of the metal–insulator transition, adding weight to the earlier findings in related two-dimensional systems. Their primary evidence consists of photoemission spectra, taken at temperatures of 12 and 140 kelvin, that simultaneously measure the energy and momentum distribution of electrons. At the lower temperature the spectra show the opening of a textbook insulating gap. The authors buttress this result with a second, surprising, finding: at 12 kelvin, the rumpling distortion found in earlier experiments⁸ disappears and a uniform flat surface returns.

The observation of both phase transitions — electronic and structural — makes the picture particularly convincing. One final experiment measuring the 'inverse photoemission' spectrum (which reveals the distribution of empty states just above the gap) would clinch the case. For this, however, we must await improvements in resolution. In the meantime, the results of Cortés and colleagues restore a sense of order to our limited, but growing, understanding of Mott insulators on surfaces. Such an understanding may help to unravel other mysteries, from high-temperature superconductivity to ultracold atoms. And as the dimensions of electronic devices continue to shrink, Mott's theory may itself become the

new cornerstone for describing how electrons behave at the very smallest scales. ■
Steven C. Erwin is at the Center for Computational Materials Science, Naval Research Laboratory, Washington DC 20375, USA.
e-mail: erwin@dave.nrl.navy.mil

1. Bloch, F. *Z. Phys.* **52**, 555–600 (1928).
2. Wilson, A. H. *Proc. R. Soc. Lond. A* **133**, 458–491 (1931).
3. Cortés, R. *et al. Phys. Rev. Lett.* **96**, 126103 (2006).
4. Mott, N. F. *Metal-Insulator Transitions* (Taylor and Francis, London, 1974).
5. Weitering, H. H. *et al. Phys. Rev. Lett.* **78**, 1331–1334 (1997).
6. Hellberg, C. S. & Erwin, S. C. *Phys. Rev. Lett.* **83**, 1003–1006 (1999).
7. Carpinelli, J. M. *et al. Nature* **381**, 398–400 (1996).
8. Carpinelli, J. M. *et al. Phys. Rev. Lett.* **79**, 2859–2862 (1997).
9. Avila, J. *et al. Phys. Rev. Lett.* **82**, 442–445 (1999).

CELL BIOLOGY

Skin care by keratins

M. Bishr Omary and Nam-On Ku

Keratin proteins perform several functions in skin cells, including those of providing mechanical support and protection against injury. But it seems they also have a more active part to play in healing wounds.

Like the musculoskeletal framework of humans or the steel-girder scaffolds of buildings, the intermediate filaments of the cytoskeleton shield cells from mechanical forms of injury^{1–3}. Intermediate filaments are made up of a large family of tissue-specific proteins, including desmin in muscle, neurofilaments in neurons and keratins in the epithelial cells that line organs such as the skin^{1–3}. Although intermediate filaments are well known for their protective properties, it seems that they may also have a role in damage repair. On page 362 of this issue, Kim *et al.*⁴ provide a direct mechanistic link between an increase in the expression of the gene for keratin K17 and the characteristic increase in protein synthesis and cell growth seen in the cells around wounds.

Wound healing involves the orchestration of numerous events in the recovering cells and those adjacent to them^{5,6}, such as cell growth, cell division and migration, and the upregulation of many genes — including those encoding several intermediate filament proteins (Fig. 1). In response to skin injury, for example, the cells surrounding the wound ramp up protein synthesis and enlarge to help seal the wound. In these cells, there is a rapid increase in the expression of several keratins (K6, K16 and K17)⁶.

Kim *et al.* began their study by pursuing their previous observation that injuries in mouse embryos lacking K17 show a striking delay in wound closure^{4,6}. They noted that the wound-edge cells from such embryos did not swell up as usual, and were about 40% smaller than those of normal embryos. Moreover,

these cells did not show the full protein-synthetic activity that typically accompanies cell growth. This initial link with cell growth led Kim *et al.* to explore the role of a regulatory enzyme called mammalian target of

rapamycin (mTOR). This enzyme can regulate cell growth and proliferation in other systems by controlling protein synthesis⁷. Kim *et al.* found that the activation of mTOR seen in normal cells is reduced in cells that lack K17.

So how does K17 affect mTOR signalling? The authors found that K17 binds to an adaptor protein called 14-3-3 σ , which belongs to the 14-3-3 protein family⁴. The 14-3-3 proteins bind to more than 100 other proteins, primarily at particular phosphoserine/phosphothreonine residues — that is, serine (Ser) and threonine (Thr) residues that have a phosphate group attached. One of their many functions is to regulate where their binding partners can reside in the cell⁸. Mutation of two sites on the K17 protein that looked likely to act as 14-3-3 binding motifs (Thr 9 and Ser 44) not only blocked the binding of K17 to 14-3-3, but also prevented the normal activation of mTOR and translocation of 14-3-3 from the nucleus to the cytoplasm in cultured cells. The investigators then went full circle by showing that reintroducing K17 into keratinocytes lacking K17 restored the movement of 14-3-3 from the nucleus to the cytoplasm, leading to stimulation of mTOR signalling and increased protein synthesis and cell size. By contrast, introducing K17 that was mutated at its 14-3-3-binding sites had no significant effect.

These findings directly implicate the keratin cytoskeleton in the regulation of protein synthesis and cell size, highlighting an overlooked non-mechanical function for skin keratins

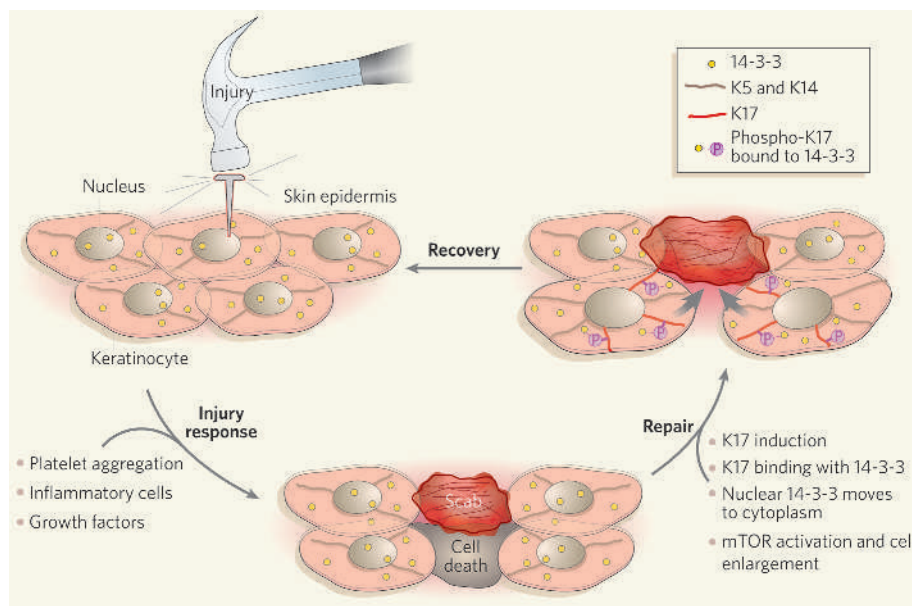


Figure 1 | The skin injury response. Injury to keratinocyte cells of the uppermost skin layer (epidermis), or injury to other tissues, triggers an elaborate repair response that includes possible bleeding and aggregation of platelet cells from the blood, scab formation, infiltration of inflammatory cells into the wound, and release of growth factors. The size, shape and adhesive properties of the keratinocytes surrounding the wound also change. These alterations result in part from changes in gene expression, including the induction of genes encoding the K6, K16 and K17 keratins. K17 probably undergoes phosphorylation, which results in binding of cytoplasmic K17 to the adaptor protein 14-3-3 and movement of nuclear 14-3-3 to the cytoplasm. As demonstrated by Kim *et al.*⁴, K17 induction and binding to 14-3-3 are key events for the activation of the mTOR signalling pathway and the ensuing stimulation of cell growth and migration.

that may extend to intermediate filament proteins in other tissues. Two other intermediate filament proteins, K18 (a major keratin in secretory epithelia)⁹ and vimentin (found in fibroblasts and endothelial cells, among others)¹⁰, are known to bind to 14-3-3. Notably, mutation of K18 at the Ser 33 site that regulates 14-3-3 binding stops the movement of 14-3-3 from the nucleus during liver regeneration⁹. Given that increased expression of intermediate filaments is a protective response to injury in many tissues^{3,6}, such upregulation of these filaments, coupled to interaction with 14-3-3, may have similar 'healing' roles in other tissues. However, intermediate filament induction will probably have many roles that may or may not require 14-3-3, as well as being mechanical and non-mechanical in nature.

The 14-3-3 proteins are already implicated in wound healing, as they can facilitate migration of keratinocyte skin cells to close the wound by regulating the adhesion molecules that hold the cells together¹¹. Presumably, this is a late event in wound healing, as K17 binding to 14-3-3 probably occurs earlier and regulates other 14-3-3 binding partners (for example, modulators of the mTOR pathway⁴) to allow increased protein synthesis and progression of wound repair (Fig. 1). The regulation, stoichiometry and precise timing of K17 binding to 14-3-3 after tissue injury remain to be determined. For instance, although Kim *et al.*⁴ examined Thr 9 and Ser 44 in K17, it is not clear whether both of these need to be phosphorylated for 14-3-3 to bind. Developing genetic models⁹, such as those in mice where K17–14-3-3 binding is blocked, and reagents (for example, antibodies specific for the phosphate tag) to track the precise timing of K17–14-3-3 binding, should further clarify this regulatory pathway.

The induction of K17 now becomes one of several established 'command posts' for the regulation of protein synthesis^{4,7}. It may well be that K17 induction is part of a regulatory loop that can, depending on context, further enhance protein synthesis and promote cell growth.

Additional questions are raised by this work. For instance, what is the significance of nuclear 14-3-3 and its redistribution during wound healing? 14-3-3 is linked to the regulation of cell division, and can affect gene expression by binding to the histone proteins that are associated with chromosomes^{8,12}. So are these processes involved in the injury response? Also, can some of the skin and nail disorders caused by mutations in the K17 gene³ be explained by signalling effects mediated by 14-3-3 proteins? K17 is induced in several conditions, including psoriasis and some cancers⁶, so what is its role in these disorders? Further areas for investigation include the possible interaction of K17 with 14-3-3 proteins other than the σ -isoform and involvement of K17–14-3-3 binding in other known functions of 14-3-3 (ref. 8). By linking the induction of

keratin in response to skin damage with signalling through 14-3-3 proteins and the consequent effects on protein synthesis and cell growth, Kim *et al.* have made an exciting advance towards solving such problems. ■

M. Bishr Omary and Nam-On Ku are in the Department of Medicine, VA Palo Alto Health Care System and Stanford University School of Medicine, 3801 Miranda Avenue, Mail code 154J, Palo Alto, California 94304, USA.
e-mail: mbishr@stanford.edu

1. Fuchs, E. & Cleveland, D. W. *Science* **279**, 514–519 (1998).
2. Herrmann, H., Hesse, M., Reichenzeller, M., Aebi, U. &

- Magin, T. M. *Int. Rev. Cytol.* **223**, 83–175 (2003).
3. Omary, M. B., Coulombe, P. A. & McLean, W. H. I. *N. Engl. J. Med.* **351**, 2087–2100 (2004).
4. Kim, S., Wong, P. & Coulombe, P. A. *Nature* **441**, 362–365 (2006).
5. Woolley, K. & Martin, P. *BioEssays* **22**, 911–919 (2000).
6. DePianto, D. & Coulombe, P. A. *Exp. Cell Res.* **301**, 68–76 (2004).
7. Hay, N. & Sonenberg, N. *Genes Dev.* **18**, 1926–1945 (2004).
8. Dougherty, M. K. & Morrison, D. K. *J. Cell Sci.* **117**, 1875–1884 (2004).
9. Ku, N. O., Michie, S., Resurreccion, E. Z., Broome, R. L. & Omary, M. B. *Proc. Natl Acad. Sci. USA* **99**, 4373–4378 (2002).
10. Tzivion, G., Luo, Z. J. & Avruch, J. *J. Biol. Chem.* **275**, 29772–29778 (2001).
11. Santaro, M. M., Gaudino, G. & Marchisio, P. C. *Dev. Cell* **5**, 257–271 (2003).
12. Macdonald, N. *et al. Mol. Cell* **20**, 199–211 (2005).

PLASMA PHYSICS

Cool vibes

Thomas C. Killian

Ultracold plasmas blur the classical boundaries between the different states of matter. Newly observed electron-density waves could become useful probes of how electrons behave in this exotic regime.

The complex soups of ions and electrons known as plasmas are ubiquitous in nature, whether in a simple flame, the aurora borealis, the Sun or elsewhere. They are also crucial for technologies such as fluorescent lighting, or for plasma processing in microelectronics. Determining how properties such as temperature and density change inside a plasma is often challenging, but it is necessary for optimizing technological applications and also for understanding fundamental physical phenomena. Some of the most valuable probes developed to determine these properties are based on collective motions of the charged particles that are sensitive to conditions in a plasma. Writing in *Physical Review Letters*, Fletcher *et al.*¹ describe observations in an ultracold plasma of what seem to be oscillations in electron density known as Tonks–Dattner resonances. The history of these resonances reads like a good detective novel, and the results imply that they may become a powerful diagnostic tool.

Students taking a first course in plasma physics learn that, in an infinite, homogeneous plasma, the electron density oscillates naturally at a frequency, ω_p , that is proportional to the square root of the electron density. This 'electron plasma oscillation' has a simple physical description: if the electron density drops because electrons are, on average, moving away from the more massive and sluggish ions, strong electric fields develop that pull the electrons back. As there is little damping for electron motion, however, the electrons overshoot, and the local electron density tends to oscillate around its equilibrium value.

In plasmas with densities that vary over a length scale greater than that of the electron

oscillations, the structure of these electron collective modes becomes richer. This fact was discovered in the 1950s, during experiments motivated by a prediction² that radio waves with a frequency of ω_p should reflect off plasma in the tail of a meteor. To general surprise, cylindrical plasmas designed to simulate meteor tails in fact scattered incident radio waves at multiple frequencies near ω_p , indicating the existence of additional collective modes besides the electron plasma oscillation expected from the theory of homogeneous plasmas.

The additional modes were later named Tonks–Dattner resonances, and their frequencies increased as the diameter of the plasma column decreased³, in a manner reminiscent of the increase in pitch from a pipe organ as the pipe becomes shorter. A quantitatively accurate model of this resonance phenomenon⁴ drew on the insight that the thermal motion of the electrons can transform the electron plasma oscillation into an extended density wave that propagates rather like a sound wave. In the experiment, the plasma density — and so ω_p — increased towards the central axis. The Tonks–Dattner modes were resonant density waves of various frequencies ω , each confined to a region at the outer edge of the plasma where their ω was greater than ω_p (Fig. 1). Farther in, at the inner radius at which $\omega = \omega_p$, the Tonks–Dattner wave decayed abruptly (became 'evanescent'), because at this point the higher-density electrons were moving quickly enough to cancel, or screen, the oscillating electric field underlying the wave. Higher-frequency resonances represented shorter-wavelength 'standing waves', for which more wavelengths of the

EVOLUTION

Experiments in botany

Amborella trichopoda, a shrub from New Caledonia in the Pacific Ocean, is believed to be the last remnant of one of the most ancient lineages of angiosperms, or flowering plants. As William E. Friedman reports in this issue (*Nature* **441**, 337–340; 2006), its red fruits, shown here, contain a relic of what seems to have been a period of intense evolutionary experimentation in early angiosperm history.

In nearly all angiosperms, the female gametophyte — egg-producing structure — is a sac containing seven cells and eight

nuclei (see Fig. 1 of the paper on page 337). Apart from the egg cell itself, which is fertilized to become the embryo, a so-called central cell is fertilized to become the supporting tissue or endosperm. These are the elements required for the phenomenon of double fertilization unique to angiosperms. Making up the numbers are three antipodal cells and two synergids — sterile cells that accompany the egg.

Amborella dares to be different, and has one extra synergid. This may seem a trifling matter, but it is akin to finding a fossil amphibian with an

extra leg: the developmental ramifications are equally important, for Friedman's detailed examination of *Amborella* development shows that the egg cell is — uniquely — a lineal sister of a synergid, rather than a mitotic sister of one of the nuclei of the central cell.

These findings add to the oddities of other relics of early angiosperm evolution, such as water lilies (Nymphaeales) whose mature female gametophytes contain only four cells. The major lineages of angiosperms are believed to have become established in an interval of around 15 million years, 130 million years ago. *Amborella* and other living fossils offer an



T. LEMIEUX, UNIV. COLORADO

immensely valuable window into how the reproductive structures of flowering plants evolved.

Henry Gee

density wave fitted between the edge of the plasma and the radius at which evanescence occurred. This explained the multiple reflections observed in the cylindrical-plasma experiment.

The experiments of Fletcher *et al.*¹ probed ultracold plasmas, which are formed from laser-cooled atoms by the ionizing action of an intense laser pulse. With electron temperatures ranging from 1 to 10^3 kelvin, and ion temperatures near 1 K, these plasmas push the envelope of neutral-plasma physics beyond the conventional range of between 10^3 K (for a flame) and 10^7 K or higher (for the solar core or a nuclear fusion reactor). At such very low temperatures, interactions between the particles can dominate the particles' random thermal motion, creating a 'strongly coupled' plasma that acts as a liquid or a solid rather than as a standard, gaseous plasma. Studying this strong coupling in the ultracold regime in a laboratory could illuminate the properties of systems that are strongly coupled as a result of their high density, for example the interiors of gas giant planets such as Jupiter, or the surfaces of neutron stars.

Ultracold plasmas are also notable for their well-defined density distribution, which drops off in a gaussian fashion in all directions radiating out from a central point, as well as their small characteristic size of about 1 mm or less. They are also formed far from their thermal equilibrium: electrons are confined in the plasma by electric attraction to the ions, and after plasma formation, the entire distribution expands into a surrounding vacuum and dissipates in about 100 microseconds.

To study electron oscillations, Fletcher *et al.*¹ applied radio waves to their ultracold plasma, just as in the meteor experiments. When the radio frequency matched the resonance of a collective mode, the applied radio field pumped energy into the electron cloud. This caused some electrons to boil out of the

plasma and escape the pull of the ions, allowing them to be counted by a charged-particle detector. At low excitation intensity, only a single resonance, matching the electron plasma oscillation, was observed.

Fields of higher intensity, however, excited

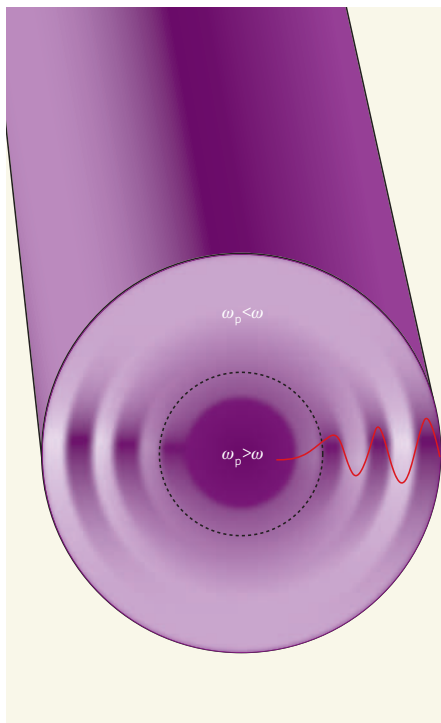


Figure 1 | All wave together. Cross-section of a Tonks–Dattner wave of frequency ω in a cylindrical plasma, such as a meteor tail. Darker colour indicates higher plasma density and the red line indicates a radial trace of the amplitude of the density variation associated with the wave. The wave is confined in the low-density region, where $\omega_p < \omega$. When $\omega_p > \omega$, electron screening quickly damps the oscillation. In ultracold plasmas, Tonks–Dattner waves are spherical, but are still confined to the outer region of low density.

a series of higher-frequency modes. As the plasma expanded, the mode frequencies decreased in accordance with what would be expected for Tonks–Dattner modes. It might seem surprising that thermal electron motion — which is crucial to the formation of such modes — would be important in ultracold plasmas. But the small plasma size means that the resonant wavelength must be small, which increases the likelihood that resonant modes form.

One difference between the observed modes and the classic Tonks–Dattner picture is that the gaussian-shaped ultracold plasmas have no hard edge like that of the confined cylindrical plasmas of the earlier experiments. How exactly a standing wave forms in this open geometry still needs to be worked out. With that theoretical input, however, the resonance frequencies should provide an accurate measure of the electron temperature, a quantity that is essential for studying the approach towards thermal equilibrium of ultracold plasmas near, or in, the strongly coupled regime.

The approach towards equilibrium is a fundamentally challenging problem, as there are many competing processes that affect the electron temperature, such as cooling through adiabatic expansion of the plasma, or heating caused by the recombination of electrons and ions to form neutral atoms. The collective modes observed by Fletcher and colleagues¹ could prove to be the best probe for investigating it.

Thomas C. Killian is in the Department of Physics and Astronomy, Rice University, Houston, Texas 77251-1892, USA.
e-mail: killian@rice.edu

1. Fletcher, R. S., Zhang, X. L. & Rolston, S. L. *Phys. Rev. Lett.* **96**, 105003 (2006).
2. Herlofson, N. *Arkiv Fysik* **3**, 247 (1951).
3. Dattner, A. *Phys. Rev. Lett.* **10**, 205–206 (1963).
4. Parker, J. V., Nickel, J. C. & Gould, R. W. *Phys. Fluids* **7**, 1489–1500 (1964).

DEVICE PHYSICS

A bug-beating diode

Asif Khan

A diode that emits light at a shorter wavelength than ever before shows huge — albeit destructive — technological promise. But further work is needed to ensure that this promise is fulfilled.

Earth's ozone layer completely blocks solar light of very low ultraviolet wavelengths. Biological organisms on Earth have therefore never developed a tolerance for this 'UVC' radiation, and artificially generated UVC light has become a useful tool in the treatment and destruction of bacteria, yeast, viruses and fungi. Mercury, xenon and deuterium lamps are currently the main sources of UVC light, but the high operating voltages and large size of such lamps, together with the environmental dangers of using mercury, preclude their use in techniques for disinfection, for air and water purification and in biomedicine. On page 325 of this issue, Taniyasu *et al.*¹ report the potentially revolutionary production of UVC light using semiconductor light-emitting diodes (LEDs) based on aluminium nitride, which could be powered by low-voltage solar cells.

An LED generally comprises a junction between semiconductor materials with two different types of conductivity, electron conductivity and hole conductivity, that are equivalent to the movement of negative and positive charge carriers, respectively. At the junction, electrons and holes recombine, emitting energy in the form of light. Highly efficient blue-green LEDs, using semiconducting aluminium indium gallium nitride (AlInGaIn), were initially reported² in the early 1990s. Thus, for the first time, miniature low-voltage light sources

in all three primary colours — red, green and blue — became available, opening up a multi-billion-dollar market for the lighting and display industries. LEDs that emit at even lower, ultraviolet wavelengths, divided into UVA, UVB and UVC, soon followed^{3,4} (Fig. 1).

The main obstacle to the development of both blue-green and ultraviolet LEDs has been the difficulty in manipulating the conduction properties of various AlInGaIn materials that possess the appropriate optical properties. The level and type of conductivity of such materials can be controlled by incorporating impurities into them such as silicon and magnesium, in a process known as doping. Indium nitride (with no aluminium or gallium) absorbs visible and ultraviolet light, and is typically a good conductor. Manipulating its conductivity by doping is also relatively easy. The same is true for gallium nitride, but this compound only absorbs ultraviolet radiation, and is completely transparent to visible light.

Typically, the wavelength of light emitted from a semiconductor LED is nearly the same as that of the light that it absorbs. And as the proportion of aluminium in the alloy increases, so does its transparency down to lower ultraviolet wavelengths. Thus, whereas an LED based on indium gallium nitride emits visible radiation, those based on aluminium gallium nitride and aluminium nitride generate ultraviolet

light with a wavelength in the UVB and the UVC portion of the solar spectrum, respectively. But unfortunately, as the aluminium fraction increases, so too does the difficulty of doping. It is hardest of all for aluminium nitride — which is, in fact, an insulator.

There are several reasons why it is difficult to turn aluminium nitride or aluminium gallium nitride with large aluminium fractions into conductors. It requires the incorporation of many impurity atoms to generate additional positive and negative current carriers, but thermal vibrational effects caused by the high-temperature conditions required for growing high-quality crystalline aluminium nitride layers work against this incorporation process. In addition, the layers are deposited over non-nitride materials, such as sapphire, generating a large number of defects that hinder doping. Defects are also generated if the number of incorporated dopant species becomes too numerous, resulting in a self-compensation effect that actually reduces conductivity. Finally, the effect of doping material on the conduction properties can be compensated by that of other materials such as hydrogen gas, used as part of the manufacturing process.

Taniyasu and colleagues' crucial contribution¹ is to overcome these obstacles by using a variation of the standard growth conditions. They control the amount of dopant precisely to avoid self-compensation, and use annealing procedures to avoid compensation effects from the reactant gases such as hydrogen. They are thus able to impart to aluminium nitride sufficient conductivity for both positive and negative current carriers. By combining layers of the two conducting types, they fabricated LEDs emitting at a UVC wavelength of 210 nanometres — at present the shortest reported wavelength of any LED when injected with electrical current.

Developments are needed in two areas to improve such UVC LEDs to the point where they can be used in devices: first, an increase in their efficiency by a factor of at least a million; and second, a reduction in their operating voltage to well below the 25 volts used by Taniyasu and colleagues¹. The first development will require substantial improvements in the crystalline quality of the aluminium nitride layers; the second will necessitate more-efficient doping to bring about a near thousandfold increase in their room-temperature conductivity. Both challenges are difficult: further bold and innovative research will be required if the promise of UVC LEDs is to be fulfilled.

Asif Khan is in the Photonics and Microelectronics Laboratory, Department of Electrical Engineering, University of South Carolina, 301 South Main Street, Columbia, South Carolina 29208, USA. e-mail: asif@engr.sc.edu

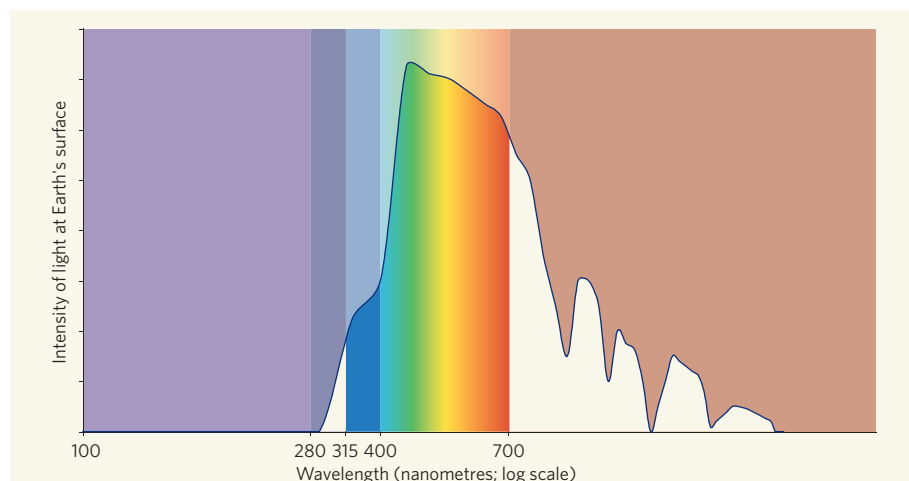


Figure 1 | Diodes in the solar radiation spectrum. The blue line indicates the approximate intensity of radiation that penetrates Earth's atmosphere at wavelengths from the ultraviolet (UV) through the visible to the infrared. Light at low UVC wavelengths is completely absorbed by the atmospheric ozone layer, so organisms on Earth have developed no tolerance to it. A small-scale, low-power light source emitting in the UVC range could therefore be useful for sterilization applications. Previous work had led to indium gallium nitride and aluminium gallium nitride light-emitting diodes (LEDs) that emit at blue² and UVB (ref. 4) wavelengths, respectively. But, until now¹, difficulties controlling the conduction properties of aluminium nitride prevented the production of LEDs at UVC wavelengths.

1. Taniyasu, Y., Kasu, M. & Makimoto, T. *Nature* **441**, 325–328 (2006).
2. Nakamura, S. *et al.* *Jpn J. Appl. Phys.* **34**, L797–L799 (1995).
3. Kinoshita, A. *et al.* *Appl. Phys. Lett.* **77**, 175–177 (2000).
4. Sun, W. *et al.* *Jpn J. Appl. Phys.* **43**, L1419–L1421 (2004).

NEWS & VIEWS FEATURE

MICROBIAL SCIENCES

The superficial life of microbes

Roberto Kolter and E. Peter Greenberg

The social activities and organization of bacteria are crucial to their ecological success. But it is only in recent years that we have begun to study these secret societies.

Most surfaces on this planet teem with microbial life, creating ecosystems of diverse organisms that flourish in slimy beds of their own making. The plaque encrusting our teeth, the slippery coating on river stones, the gunge clogging up water pipes or infected wounds: these are just a few examples of the microbial 'biofilms' that form anywhere there is a surface with a little moisture and some nutrients. Although microbes by and large live in such biofilm communities, most of our understanding of their physiology stems from experiments using liquid cultures of dispersed, free-swimming 'planktonic' cells. In the past decade, however, the number of studies performed on surface-associated microbes has increased dramatically. Today, we recognize that most, if not all, microbial species can form biofilms. The physiological differences between free-living individuals and communal biofilm-associated cells are becoming apparent, as are the regulatory mechanisms that underlie the switch between these two lifestyles.

The diverse universe of biofilms

When individual bacterial cells encounter a surface under conditions propitious for growth, they almost invariably undergo dramatic lifestyle changes. The nomadic single cells settle down on the surface, where they divide

and establish a sedentary yet remarkably diverse community (Fig. 1a). These are communities in the sense that we humans organize ourselves into communities with division of labour — as the surface-associated population grows, the biofilm becomes increasingly sophisticated in its activities, with individual cells taking on specific tasks. As a result, biofilms can develop intricate architectures; striking mushroom-like structures can bloom on submerged surfaces, and aerial projections sprout from surfaces exposed to the air (Fig. 1b,c). The diversity of biofilm architectures is akin to a miniature coral reef, with their structures differing enormously depending on the species present and the environmental conditions.

Natural biofilms nearly always harbour a multitude of microbial species — the region around a single tooth, for example, is often sheathed in a community of several hundred species¹. Within such microbial ecosystems, species richness may ensure stability of the ecosystem in the face of changing environmental conditions, very much as species richness is thought to aid the stability of macro-scale ecosystems such as a tropical rainforest.

Even in the context of artificial, single-species biofilms, strain diversification seems to be the rule. Unlike the growth of bacteria in

liquid cultures, which produce homogeneous populations of genetically identical cells, growth in biofilms generates a large amount of genetic diversity². How can a single cell, with a single genetic complement, give rise to a biofilm population in which the individual cells are genetically different from one another? The simplest explanation may be that in any biofilm, individual cells are stuck in the same place, attached to their neighbours and the slime that surrounds them, so their access to nutrients will vary as gradients form within the biofilms through metabolic activity. As a consequence, many microniches are likely to arise, as are random spontaneous mutants that can exploit those microniches. As various mutants grow in their location, the result will be a biofilm containing a multitude of genotypes. This genetic diversity within a strain may provide a sort of insurance, as the population can adapt better to sudden environmental changes than can a genetically homogeneous population.

To understand these minute ecosystems, it is necessary to be able to grow, observe and manipulate biofilms in the laboratory. This is accomplished using transparent flow cells, where bacteria attach to a glass surface and are continuously fed fresh nutrients³. The ensuing biofilm growth can be followed using confocal laser microscopy. In such flow cells,

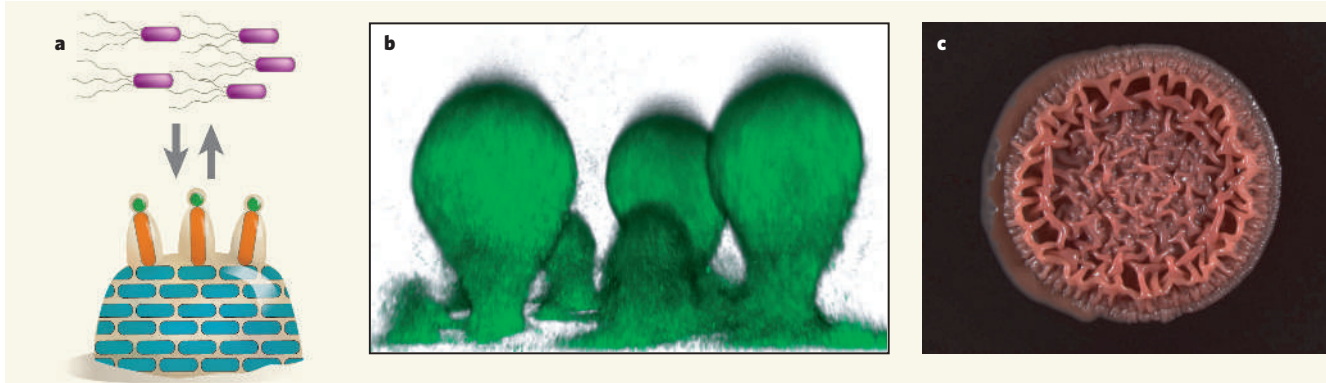


Figure 1 | Biofilm formation and architecture. **a**, Individual bacteria undergo a reversible lifestyle switch between a nomadic and a sedentary existence. In the process, they lose motility and become enclosed in a gooey extracellular matrix. As the community grows, different cell types appear as the original strain diversifies to allow cells to take on different tasks or to exist in different 'microniches', shown by differently coloured cells.

b, These mushroom-like structures are characteristic of many submerged biofilms. In this *Pseudomonas aeruginosa* biofilm, which was grown in a flow cell, they are about 150 μm high. **c**, A biofilm of *P. aeruginosa* grown on a semi-solid agar surface (about 10 mm in diameter). The colony displays aerial projections or wrinkling characteristic of an air-exposed biofilm.

experimental reproducibility is best obtained using single-species systems. As a consequence, many microbial species have been analysed as monocultures. Yet, in natural settings, these microbes nearly always grow in the context of multi-species communities. To approximate the natural setting more closely, investigators have begun to foray into systems containing two or more microbial species. Some potentially interesting multi-species model systems are poised for molecular analyses. Among these are mixed-species biofilms of the opportunistic pathogens *Pseudomonas aeruginosa* (a bacterium) and *Candida albicans* (a fungus)⁴; two bacterial species that colonize teeth, *Streptococcus gordonii* and *Veillonella atypica*⁵; and two species that occur in the soil environment around plant roots, *P. aeruginosa* and *Agrobacterium tumefaciens*⁶.

Sticking together

Given the diversity of species that form biofilms, it is not surprising that the molecular mechanisms involved in biofilm development reflect the remarkable variety of the microbial world. Different species build biofilms differently, and many strains have multiple biofilm-formation pathways. For instance, some mutants of the soil bacterium *Pseudomonas fluorescens* that cannot make a biofilm when grown with glucose as the sole carbon source do form biofilms when the sole carbon source is glutamate⁷.

So, do any general principles emerge? One feature that does seem to be common to all biofilms is that the cells secrete a 'matrix' to hold themselves in place and to provide a buffer against the environment. The make-up of each matrix is different, and, depending on the contributing species and environmental conditions, there can be various mixes of polysaccharides, proteins and even nucleic acids⁸.

The rich diversity among the matrix components can be glimpsed by mentioning what is known about the polysaccharides secreted by just one species, *P. aeruginosa*, one of the most-studied organisms in terms of biofilm formation. Many strains of this bacterium have the capacity to extrude at least three types of extracellular polysaccharide, called alginate, Pel and Psl. Alginate was initially considered the major biofilm polysaccharide, but it now seems that it is a major matrix component only of the biofilms formed by unusual variants⁹. Pel and Psl are apparently more ubiquitous: the genes involved in the synthesis of Pel occur in all *P. aeruginosa* strains studied, whereas the genes associated with Psl are present in only some^{10–12}. But even the Pel genes have diverse regulatory regions, suggesting that the various strains express this matrix component differently.

Lifestyle choices

Motile bacteria generally have flagella — corkscrew-like appendages that rotate to propel them. These structures occur only in the microbes' planktonic form, and it now seems

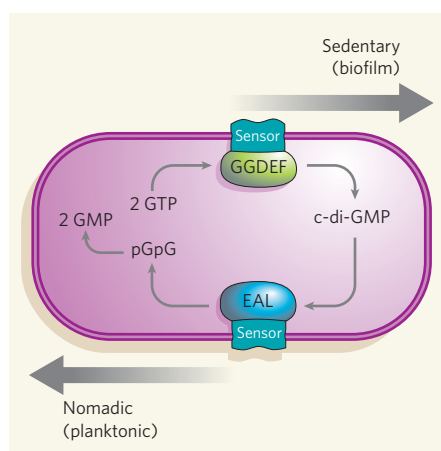


Figure 2 | The lifestyle switch. The second-messenger molecule c-di-GMP mediates the switch between nomadic and sedentary lifestyles. Many species of bacteria have a multitude of enzymes capable of making c-di-GMP from GTP (GGDEF proteins), and of breaking it down by hydrolysing it into GMP (EAL proteins). Overall, it seems that increased intracellular levels of c-di-GMP favour a sedentary existence, whereas reduced levels of the second messenger favour a nomadic existence.

that production of flagella and of extracellular matrix are mutually exclusive processes: bacteria give up the ability to move in order to settle down. The regulatory circuitry involved in this lifestyle switch is beginning to be mapped out. In *Bacillus subtilis*, for instance, the switch between flagellar synthesis and matrix production is controlled by the master gene regulator SinR (refs 13–15). This protein directly represses genes encoding matrix components and activates those encoding flagellar components (seemingly indirectly). SinR activity itself is antagonized by another protein, SinI, whose abundance and activity are modulated by changing environmental conditions. The inverse regulation of genes involved in motility and matrix components is apparent in several other organisms, including *E. coli*, *P. aeruginosa*, *Salmonella enterica* and *Vibrio cholerae*, although the specific regulatory molecules used have not been identified^{16,17}.

But how is the lifestyle switch thrown? An individual cell must have some way of recognizing that it is near a suitable surface, and of passing that information on to the master gene regulators and other effectors so that it can cease roaming and settle down. In several species, the messenger seems to be a small cytoplasmic molecule called bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP)¹⁸ (Fig. 2). Low intracellular c-di-GMP concentrations are found in planktonic cells, but these levels rise once the cells have given up their nomadic lifestyle. Furthermore, it looks as though c-di-GMP can control the synthesis of diverse cellular components involved in both motility and matrix production in response to changing environments.

The synthesis and degradation of c-di-GMP

are carried out respectively by proteins that have structural domains possessing diguanylate cyclase and phosphodiesterase enzymatic activities (referred to as GGDEF and EAL domains, respectively, because of the highly conserved amino-acid sequences they contain). Genomic analyses of dozens of microbial genomes show that most bacteria harbour a multitude of these proteins — *Vibrio vulnificus*, an extreme example, has 66 GGDEF proteins and 33 EAL proteins¹⁸. Often, the GGDEF or EAL segment is fused to one of several types of environment-sensing domains, implying that the production and breakdown of c-di-GMP are controlled by the bacterium's surroundings.

The GGDEF and EAL proteins, and by extension c-di-GMP, are involved in many diverse cellular activities, but the mechanism by which c-di-GMP acts remains largely unknown. It may act to integrate the many external inputs sensed by GGDEF and EAL proteins, allowing this single messenger to instigate multiple indirect effects. Or perhaps there are small cellular 'compartments' where local fluctuations in the concentration of c-di-GMP cause independent outcomes. Initially, c-di-GMP was shown to function by activating the enzyme that synthesizes extracellular cellulose in *Gluconoacetobacter xylinus*. This activity occurs at what is known as the post-translational level, once the target protein, in this case the enzyme, has been synthesized¹⁹. By extension, investigators believed c-di-GMP acted exclusively at the post-translational level on enzymes in other bacteria. But it now seems that c-di-GMP can affect many processes, not only by acting post-translationally, but also by regulating gene expression. For example, it functions at the level of gene-transcription control in *P. aeruginosa*²⁰. No gene-regulatory proteins that bind to c-di-GMP have yet been identified conclusively, so how this messenger passes on the external signal to the master regulators is unclear.

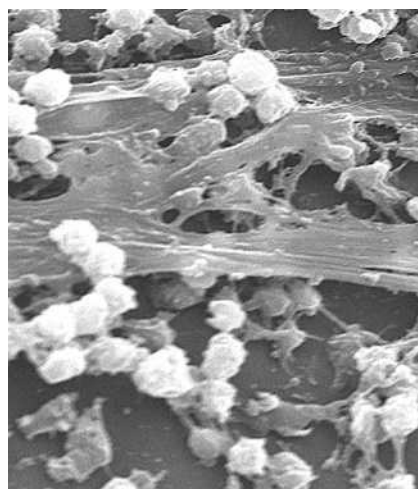
Even though the structure of c-di-GMP and its involvement in regulating cellulose synthesis in *G. xylinus* were reported²¹ in 1987, it was only in the past few years, when it became apparent that c-di-GMP occurs in bacterial pathogens, that interest in this molecule really intensified. In this sense, the history of c-di-GMP research resembles the history of quorum sensing — the means by which bacteria detect the presence of others of their kind. As early as 1968, there were reports of what we now know as quorum sensing in the marine bacterium *Vibrio fischeri*²². But it was only in the mid-1990s that quorum sensing was recognized as a feature of many pathogens²³, leading to a flood of interest in this process. One has to wonder how many interesting molecular mechanisms have already been found in microbes but have yet to receive much attention because they are not being studied in pathogens.

Box 1 Biofilms in human health and disease

Environmental microbiologists and engineers have long recognized that biofilms form on surfaces — clogging pipework, for instance, or aiding water clean-up — but this fact has only recently been appreciated fully by those in the medical sciences. The intimate relationship between the human body and its resident microbes is now beginning to be elucidated.

Such biofilms are mostly beneficial. By colonizing our skin, teeth and the mucosal surfaces lining our gut and airways, thousands of different microbial species play an essential role in our nutrition and serve as a primary line of defence against invading pathogens. But although they are vital to our well-being, we know very little about the microbiota that thrive on and in our bodies. We scarcely know how to cultivate even a small percentage of these microbial inhabitants so as to be able to study them closely.

Some biofilms, however, are extremely harmful to us. Notably, foreign objects such as catheters and prostheses provide enticing surfaces for colonization, and microbes often quickly take up residence (for example, the image shows a *Staphylococcus aureus* biofilm found on an in-dwelling catheter, magnified 1,180 times). The resulting biofilms can become reservoirs for



CDC/R. M. DONLAN, J. CARR

systemic infection. Unfortunately, for reasons that remain poorly understood, biofilm-associated microbes are particularly impervious to many antimicrobial agents, so biofilm-related infections are difficult to treat. A major goal in this field is to develop therapeutic strategies that can control the undesirable growth of biofilms while leaving beneficial biofilms intact.

R.K. & E.P.G.

Community relations

Once an advance party of bacteria has set up camp on a pristine surface, the cells can let each other know they are there by exuding quorum-sensing signals²³. Planktonic bacteria use these chemical signals to regulate a variety of processes, depending on population density. But in many bacterial species the same signals are now known to be cues for biofilm development. This was first shown in *P. aeruginosa*, in which quorum-sensing signals control the expression of hundreds of genes throughout the genome^{24,25}. In certain surroundings, mutants defective in quorum signalling are also defective in biofilm formation; for instance, they do not grow mushroom-like structures under conditions where these structures would normally form. The mutants still make biofilms, however, and in certain conditions where mushroom structures would not usually form, these biofilms are indistinguishable from those formed by unmutated strains. This observation may be due to the fact that the effects of the quorum-sensing mutations on biofilm formation are not apparent visually, or that quorum sensing is involved in specific steps in biofilm formation such as mushroom building. So it seems that biofilm formation is a social activity that is also governed by both genetic and environmental factors (reviewed in ref. 26).

An area that will see increasing attention involves the question of whether signalling in biofilm biology occurs between species. Multi-species biofilm model systems are now available to address such questions, and there is

mounting evidence that bacteria do have systems that monitor and respond to quorum-sensing signals from other species. In some cases these systems are very specific, and in others there may be methods for sensing signals or cues produced by almost any member of the local microflora²⁷.

If different species can sense one another, how do they interact? Can biofilm communities encourage or discourage individuals in the planktonic community over entry to the biofilm? In *P. aeruginosa* biofilms, the quorum-sensing response results in the activation of a battery of potential defence systems, including a system for producing cyanide and one that releases antibiotics to attack other bacteria^{25,28}. Whether this response does indeed constitute a defence against interlopers is being studied experimentally by tagging the biofilm bacteria and the planktonic bacteria with differently coloured fluorescent proteins (G. J. Balzer, M. H. Hentzer and M. R. Parsek, personal communication).

Outlook for the future

Surfaces afford a greater capacity for organization than do liquids, and so microbes encountering surfaces can form aggregates that begin to display some attributes of multicellularity. These groups of cells are held in place by an extracellular matrix and can use intercellular signalling for communication. In doing so, they develop intriguing features that we are only just beginning to understand.

Traditionally, microbes have been studied in their planktonic forms. Only in the past

decade have the tools of molecular biology been brought to bear on biofilms — a form that has immense biological significance. With recent technological developments that allow analysis of multi-species biofilms, the next decade should see further integration of molecular approaches into the study of biofilm biology, and significant progress in our understanding of more complex problems of bacterial social behaviour. At an applied level, we can hope to develop agents that control the biology of biofilms — to allow them to be analysed further, to treat biofilm infections, and to control biofilms for beneficial uses (Box 1). Clearly, much remains to be discovered about the diverse world of bacterial biofilms, and it seems the time is ripe for such discoveries.

Roberto Kolter is in the Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

e-mail: rkolter@hms.harvard.edu

E. Peter Greenberg is in the Department of Microbiology, University of Washington, Seattle, Washington 98195, USA.

e-mail: epgreen@u.washington.edu

1. Paster, B. J. *et al.* *J. Bacteriol.* **183**, 3770–3783 (2001).
2. Boles, B. R., Thoendel, M. & Singh, P. K. *Proc. Natl Acad. Sci. USA* **101**, 16630–16635 (2004).
3. Palmer, R. J. *Jr Methods Enzymol.* **310**, 160–166 (1999).
4. Hogan, D. A., Vik, A. & Kolter, R. A. *Mol. Microbiol.* **54**, 1212–1223 (2004).
5. Eglund, P. G., Palmer, R. J. Jr & Kolenbrander, P. E. *Proc. Natl Acad. Sci. USA* **101**, 16917–16922 (2004).
6. An, D., Danhorn, T., Fuqua, W. C. & Parsek, M. R. *Proc. Natl Acad. Sci. USA* **103**, 3828–3833 (2006).
7. O'Toole, G. A. & Kolter, R. *Mol. Microbiol.* **28**, 449–461 (1998).
8. Branda, S. S., Vik, S., Friedman, L. & Kolter, R. *Trends Microbiol.* **13**, 20–26 (2005).
9. Wozniak, D. J. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 7907–7912 (2003).
10. Matsukawa, M. & Greenberg, E. P. *J. Bacteriol.* **186**, 4449–4456 (2004).
11. Friedman, L. & Kolter, R. *J. Bacteriol.* **186**, 4457–4465 (2004).
12. Jackson, K. D., Starkey, M., Kremer, S., Parsek, M. R. & Wozniak, D. J. *J. Bacteriol.* **186**, 4466–4475 (2004).
13. Kearns, D. B., Chu, F., Branda, S. S., Kolter, R. & Losick, R. *Mol. Microbiol.* **55**, 739–749 (2005).
14. Chu, F., Kearns, D. B., Branda, S. S., Kolter, R. & Losick, R. *Mol. Microbiol.* **59**, 1216–1228 (2006).
15. Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. *Mol. Microbiol.* **59**, 1229–1238 (2006).
16. Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. *Mol. Microbiol.* **53**, 1123–1134 (2004).
17. Watnick, P. I., Lauriano, C. M., Klose, K. E., Croal, L. & Kolter, R. *Mol. Microbiol.* **39**, 223–235 (2001).
18. Romling, U., Gomelsky, M. & Galperin, M. Y. *Mol. Microbiol.* **57**, 629–639 (2005).
19. Weinhouse, H. *et al.* *FEBS Lett.* **416**, 207–211 (1997).
20. Hickman, J. W., Tifrea, D. F. & Harwood, C. S. *Proc. Natl Acad. Sci. USA* **102**, 14422–14427 (2005).
21. Ross, P. *et al.* *Nature* **325**, 279–281 (1987).
22. Kempner, E. S. & Hanson, F. E. *J. Bacteriol.* **95**, 975–979 (1968).
23. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. *J. Bacteriol.* **176**, 269–275 (1994).
24. Davies, D. G. *et al.* *Science* **280**, 295–298 (1998).
25. Whiteley, M., Lee, K. M. & Greenberg, E. P. *Proc. Natl Acad. Sci. USA* **96**, 13904–13909 (1999).
26. Parsek, M. R. & Greenberg, E. P. *Trends Microbiol.* **13**, 27–33 (2005).
27. Waters, C. M. & Bassler, B. L. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).
28. Mashburn, L. M. & Whiteley, M. *Nature* **437**, 422–425 (2005).

BRIEF COMMUNICATIONS

Semantic combinations in primate calls

Putty-nosed monkeys rely on two basic calling sounds to construct a message of utmost urgency.

Syntax sets human language apart from other natural communication systems, although its evolutionary origins are obscure¹. Here we show that free-ranging putty-nosed monkeys combine two vocalizations into different call sequences that are linked to specific external events, such as the presence of a predator and the imminent movement of the group. Our findings indicate that non-human primates can combine calls into higher-order sequences that have a particular meaning.

Like most forest guenons, male putty-nosed monkeys (*Cercopithecus nictitans*) produce two acoustically distinct loud calls ('pyows' and 'hacks') in response to a range of disturbances^{2,3}. Males also call spontaneously, especially during morning foraging and evening travel to sleeping sites. In addition, these calls can function as alarm calls to warn the group of an approaching predator and to discourage attack³⁻⁵: pyows are used primarily when a leopard (*Panthera pardus*) is in the vicinity, and hacks are produced mainly in response to crowned eagles (*Stephanoaetus coronatus*) (audio files are in supplementary information).

Apart from these predator-specific call sequences, we found that males regularly combine the same two calls into a third structure, a 'pyow-hack' (or P-H) sequence, which usually consists of one, two or three pyows followed by up to four hacks (total mean duration 6.11 ± 3.02 s). These P-H sequences were emitted in response both to eagles and to leopards, and were either isolated or interspersed with other sequence types. We noticed that the group of monkeys would begin to move soon after hearing a P-H sequence.

To test whether the function of the P-H call sequence could be to instigate this movement, we carried out a field experiment in which leopard growls were played to 17 different monkey groups, each consisting of several adult females with their offspring and a single adult male, to elicit calling from the adult male (for details of methods, see supplementary information). In nine groups (52.9%), the male produced call series containing at least one P-H sequence; in the other eight groups, the males also called, but no P-H sequences were produced.

Twenty minutes later, we played a second stimulus, a series of hacks from the same position to imitate the alarm call produced by monkeys in the presence of a crowned eagle. This reliably elicited further vocal responses from all 17 males, which enabled us to relocate the

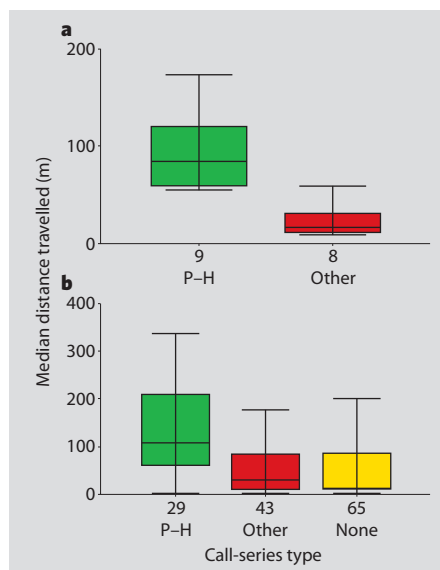


Figure 1 | Relationship between the occurrence of sequences in male call series and the distance travelled in response to the calls by groups of putty-nosed monkeys. **a**, Call series experimentally elicited by leopard growling: the group travelled significantly farther after hearing call series that contained at least one pyow-hack (P-H) sequence ($n = 9$) compared with the distance moved in response to all other sequences ($n = 8$). **b**, Naturally occurring call series: the habituated study group travelled significantly farther following call series containing at least one P-H sequence ('P-H'; $n = 29$) compared with the distance moved in response to other series that included no P-H sequences ('other'; $n = 43$) or when there were no calls made by the male ('none'; $n = 65$). Box plots indicate medians, interquartiles and ranges.

groups precisely. Although putty-nosed monkeys sometimes move through the canopy to escape from an approaching leopard, this strategy is avoided when threatened by large raptors as it would increase the risk of attack: consequently, warning hacks on their own do not trigger any movement of the group (Fig. 1).

Using a global positioning satellite unit (see methods in supplementary information), we found that the groups whose males had produced P-H sequences in response to growls had travelled significantly farther than other groups (Fig. 1a; call series containing at least one P-H sequence: median 85 m; all other series: median 17 m; two-tailed Mann-Whitney U -test, $n_1 = 9$, $n_2 = 8$, $Z = -3.27$, $P = 0.0003$).

The monkeys' response to P-H sequences

was not confined to the predator context, but was generally related to whether the group moved. We recorded a total of 72 natural call sequences (that is, not experimentally induced) over two months from the single male of a group habituated to human observers and monitored the group's travelling patterns. A large proportion of the calls contained P-H sequences (40.3%) and elicited movement of the group over significantly greater distances than after P-H-free call series (Fig. 1b; for series containing at least one P-H sequence: median 110 m; all other series: median 30 m; no calls: median 14 m; Kruskal-Wallis analysis of variance, $n_1 = 29$, $n_2 = 43$, $n_3 = 65$, d.f. = 2, $\chi^2 = 19.27$, $P = 0.00007$; P-H vs other: $Z = -3.42$, $P = 0.0005$; P-H vs none: $Z = -4.23$, $P = 0.0001$; other vs none: $Z = -1.01$, $P = 0.314$; two-tailed Mann-Whitney U -tests, Bonferroni corrections for multiple comparisons).

Most animals have a restricted repertoire of calls, with innate and structurally fixed vocalizations. Combining existing calls into meaningful sequences increases the variety of messages that can be generated⁶⁻¹⁰. The simple system used by putty-nosed monkeys encodes the presence of different types of danger and triggers group movement with just two basic call types.

Kate Arnold, Klaus Zuberbühler

School of Psychology, University of St Andrews, St Andrews KY16 9JP, UK
e-mail: kz3@st-and.ac.uk

1. Hauser, M. D., Chomsky, N. & Fitch, W. T. *Science* **298**, 1569-1579 (2002).
2. Gautier, J. P. & Gautier-Hion, A. in *How Animals Communicate* (ed. Sebeok, T.) 890-964 (Indiana Univ. Press, Bloomington, 1977).
3. Eckardt, W. & Zuberbühler, K. *Behav. Ecol.* **15**, 400-411 (2004).
4. Arnold, K. & Zuberbühler, K. *Anim. Behav.* (in the press).
5. Pohlner, Y., Arnold, K. & Zuberbühler, K. *Behav. Ecol. Sociobiol.* (submitted).
6. Cleveland, J. & Snowdon, C. T. *J. Tierpsychol.* **58**, 231-270 (1982).
7. Robinson, J. G. *Behaviour* **90**, 46-79 (1984).
8. Zuberbühler, K. *Anim. Behav.* **63**, 293-299 (2002).
9. Mitani, J. C. & Marler, P. *Behaviour* **106**, 20-45 (1989).
10. Marler, P., Evans, C. S. & Hauser, M. D. in *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (eds Papoušek, H., Jurgens, U. & Papoušek, M.) 66-86 (Cambridge Univ. Press, Cambridge, 1992).

Supplementary information accompanies this communication on Nature's website.

Received 13 December 2005; accepted 6 April 2006.

Competing financial interests: declared none.

doi:10.1038/441303a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

IMMUNOLOGY

Toll-like receptors and antibody responses

Arising from: C. Pasare & R. Medzhitov *Nature* **438**, 364–368 (2005).

Microbial components, such as lipopolysaccharides, augment immune responses by activating Toll-like receptors (TLRs). Some have interpreted this to mean that TLR signalling might not only help to initiate the adaptive immune response, but may also be required for it¹. The expanded view is shared by Pasare and Medzhitov², who conclude from an analysis of mice deficient in MyD88 (a TLR-signalling adaptor protein) that the generation of T-dependent antigen-specific antibody responses requires activation of TLRs in B cells. However, we show here that robust antibody responses can be elicited even in the absence of TLR signals. This appreciable TLR-independence of immune responses should be taken into account in the rational design of immunogenic and toleragenic vaccines.

In mice that lack both MyD88 and TRIF (another TLR-signalling adaptor protein), all known types of TLR signalling are absent³. We find that preimmune serum immunoglobulin concentrations are substantial in these mice, and that T-cell-dependent antibody responses

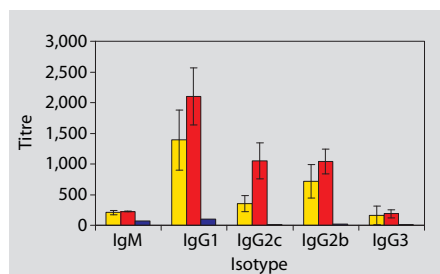


Figure 1 | Antibody responses. T-cell-dependent antibody responses in wild-type C57BL/6 mice (yellow) and in MyD88^{-/-}/TRIF^{Lps2/Lps2} double-knockout mice (red), which lack Toll-like-receptor activity. Young adult mice were given 100 µg endotoxin-free trinitrophenol-haemocyanin conjugates in alum. At day 7 post-immunization, serum antibody responses of the antibody isotypes IgM, IgG1, IgG2c, IgG2b and IgG3 were quantified by using an enzyme-linked immunosorbent assay, with titres representing the dilution giving half-maximal optical density. Means and standard errors are shown for six mice per group. Sera from unimmunized C57BL/6 mice (blue) establish the background of the assay.

to an antigen administered in alum are unimpaired (Fig. 1). It is notable that immunization with an antigen in alum, one of the few adjuvants approved for human use, promotes a TLR-independent antibody response. Further, we find that MyD88/TRIF double-deficient mice generate a robust antibody response to the T-independent antigen trinitrophenol-Ficoll (results not shown).

In the final analysis, we believe that Pasare and Medzhitov have simply demonstrated that TLR-dependent B-cell responses depend upon TLR signalling. But TLR signalling is clearly not essential for a response to typical vaccination.

D. Nemazee, A. Gavin, K. Hoebe, B. Beutler
Department of Immunology, The Scripps Research Institute, La Jolla, California 92037, USA
e-mail: nemazee@scripps.edu

1. Medzhitov, R. & Janeway, C. A. Jr *Science* **296**, 298–300 (2002).
2. Pasare, C. & Medzhitov, R. *Nature* **438**, 364–368 (2005).
3. Hoebe, K. *et al.* *Nature* **424**, 743–748 (2003).

doi:10.1038/nature04875

IMMUNOLOGY

Pasare & Medzhitov reply

Replying to: D. Nemazee, A. Gavin, K. Hoebe & B. Beutler *Nature* **441**,
doi:10.1038/nature04875 (2006)

Using a range of different experimental approaches (and not just analyses of MyD88-deficient mice), we showed that, contrary to the prevailing view, a T-helper response is not sufficient for the generation of optimal T-dependent antibody responses, and that Toll-like-receptor (TLR) signalling in B cells is also required¹. Similar results have since been obtained in human B cells². Nemazee *et al.*³ now question our conclusions¹.

We showed that the requirement for TLR signalling in antibody responses depends on the antibody isotype¹: the IgM and IgG1 isotypes are largely, but not completely, TLR-dependent; IgG2 isotypes are entirely TLR-dependent; and IgE and IgA responses

are TLR-independent. But in any event, our main discovery was that TLR signalling has to happen in B cells in order for the optimum T-cell-dependent antibody response to be generated. It was not, as claimed by Nemazee *et al.*³, simply that TLR-dependent B-cell responses depend on TLR signalling.

Figure 1 of Nemazee *et al.*³ shows responses obtained under suboptimal conditions, without essential comparisons and controls¹: for example, their antibody responses would seem negligible if compared with the response induced in the presence of TLR ligands. The results do not therefore present a meaningful counterargument to our conclusions.

Nemazee *et al.* claim that TLR signalling is

not essential for responses to “typical vaccination”, which very much depends on what is meant by ‘typical’. Most successful vaccines, such as attenuated pathogens, certainly do trigger TLR signalling, which is one of the reasons why they are successful. The subunit vaccines, on the other hand, generally do not work; the few that do, do so because they either activate TLRs or some TLR-equivalent component of the innate immune system.

C. Pasare, R. Medzhitov
Howard Hughes Medical Institute and Section of Immunology, Yale University School of Medicine, New Haven, Connecticut 06510, USA
e-mail: medzhitov@yale.edu

1. Pasare, C. & Medzhitov, R. *Nature* **438**, 364–368 (2005).
2. Ruprecht, C. R. & Lanzavecchia, A. *Eur. J. Immunol.* **36**, 810–816 (2006).
3. Nemazee, D., Gavin, A., Hoebe, K. & Beutler, B. *Nature* **441**, doi:10.1038/nature04875 (2006).

doi:10.1038/nature04876

An extrasolar planetary system with three Neptune-mass planets

Christophe Lovis¹, Michel Mayor¹, Francesco Pepe¹, Yann Alibert², Willy Benz², François Bouchy^{3,4}, Alexandre C. M. Correia⁵, Jacques Laskar⁶, Christoph Mordasini², Didier Queloz¹, Nuno C. Santos^{1,7,8}, Stéphane Udry¹, Jean-Loup Bertaux⁹ & Jean-Pierre Sivan¹⁰

Over the past two years, the search for low-mass extrasolar planets has led to the detection of seven so-called 'hot Neptunes' or 'super-Earths' around Sun-like stars. These planets have masses 5–20 times larger than the Earth and are mainly found on close-in orbits with periods of 2–15 days. Here we report a system of three Neptune-mass planets with periods of 8.67, 31.6 and 197 days, orbiting the nearby star HD 69830. This star was already known to show an infrared excess possibly caused by an asteroid belt within 1 AU (the Sun–Earth distance). Simulations show that the system is in a dynamically stable configuration. Theoretical calculations favour a mainly rocky composition for both inner planets, while the outer planet probably has a significant gaseous envelope surrounding its rocky/icy core; the outer planet orbits within the habitable zone of this star.

Since the discovery of the first extrasolar planet around a solar-type star ten years ago¹, new detections have been regularly reported by several teams, bringing the number of known extrasolar planets to more than 170 today^{2,3}. The vast majority of these discoveries have been made using the radial velocity technique, that is, the measurement of tiny radial velocity variations of the central star due to the gravitational pull of orbiting planets. This technique, intrinsically sensitive to massive, Jupiter-like planets, has been continuously improved to reach an accuracy of $\sim 1 \text{ m s}^{-1}$, leading to the discovery of planets lighter than Neptune on close-in orbits. The low end of the planetary mass distribution is now accessible to radial velocity surveys, whose ultimate detection limits have yet to be established. The accumulation of high-precision radial velocity measurements allows us to continuously refine the orbital parameters of the planets known at present and often reveals the presence of other bodies in the systems. The 17 multi-planet systems detected to date have been the subject of numerous researches studying their formation, dynamical evolution and long-term stability. They show an impressive diversity in planetary masses, orbital distances and dynamical structure, but they all share the common property of being dominated by one or more gaseous giant planets in the Jupiter-mass range. In this Article we present the first observed multiple planetary system composed only of Neptune-mass objects, orbiting the star HD 69830.

Properties of HD 69830

HD 69830 is a nearby star located 12.6 pc away from the Sun towards the southern constellation Puppis. It has spectral type K0V and visual magnitude $V = 5.95$ (ref. 4), making it just visible to the naked eye. In order to determine its basic physical properties, we performed a spectroscopic analysis based on models of stellar atmospheres⁵. We obtain an effective temperature $T_{\text{eff}} = 5,385 \pm 20 \text{ K}$ and a metallicity $[\text{Fe}/\text{H}] = -0.05 \pm 0.02$ (that is, 89% of the solar heavy element

concentration). From these parameters, and using the appropriate bolometric correction, we derive a total luminosity of $0.60 \pm 0.03 L_{\odot}$ (where $L_{\odot}$ is the solar luminosity). By interpolating within grids of theoretical stellar evolution tracks^{6,7}, we find a stellar mass of $0.86 \pm 0.03 M_{\odot}$ (where $M_{\odot}$ is the solar mass) and an age of $\sim 4\text{--}10 \text{ Gyr}$. HD 69830 is therefore an old main-sequence star, slightly less massive than the Sun.

This star has recently been under close scrutiny owing to the detection by the Spitzer Space Telescope of a strong infrared excess relative to the stellar photosphere⁸. It is attributed to emission by small grains of crystalline silicates with a size below $\sim 1 \mu\text{m}$. The grains have a temperature of $\sim 400 \text{ K}$ and must therefore be located close to the star, most probably within 1 astronomical unit (1 AU). These observations are interpreted as the signature of a massive asteroid belt within 1 AU, in which collisional processes continuously replenish the debris disk. Alternatively, the infrared emission might be caused by an evaporating super-comet recently captured onto a close orbit around HD 69830, although this scenario seems less likely.

High-precision radial velocities

We have obtained high-precision radial velocity measurements of HD 69830 during the past two years with the HARPS instrument installed on the European Southern Observatory 3.6-m telescope at La Silla Observatory, Chile. HARPS is a high-resolution ($R = 110,000$) cross-dispersed echelle spectrograph designed to achieve the highest possible radial velocity accuracy^{9,10}. It has demonstrated a long-term precision of $\sim 1 \text{ m s}^{-1}$, thereby becoming the most powerful instrument with which to detect extrasolar planets with the radial velocity technique^{11–13}. HD 69830 is a member of the high-precision sample of nearby stars that we are following closely in order to detect very-low-mass planets. We have obtained 74 data points

¹Observatoire de Genève, 51 ch. des Maillettes, 1290 Sauverny, Switzerland. ²Physikalisches Institut der Universität Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. ³Observatoire de Haute-Provence, 04870 St Michel l'Observatoire, France. ⁴Institut d'Astrophysique de Paris, 98bis Bd Arago, 75014 Paris, France. ⁵Departamento de Física da Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal. ⁶Astronomie et Systèmes Dynamiques, IMCCE-CNRS UMR 8028, 77 Av. Denfert-Rochereau, 75014 Paris, France. ⁷Centro de Astronomia e Astrofísica da Universidade de Lisboa, Observatório Astronómico de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal. ⁸Centro de Geofísica de Évora, Colégio Luis Verney, Rua Romão Ramalho, 59, 7002-554 Évora, Portugal. ⁹Service d'Aéronomie du CNRS, BP 3, 91371 Verrières-le-Buisson, France. ¹⁰Laboratoire d'Astrophysique de Marseille, Traverse du Siphon, 13013 Marseille, France.

spanning about 800 days between October 2003 and January 2006 (see Supplementary Information).

The radial velocities have been derived from the extracted spectra by the usual cross-correlation technique with a stellar template, coupled to high-precision wavelength calibration¹⁴. To estimate the uncertainties on the data points, we quadratically add the photon noise, the guiding error, the wavelength calibration uncertainty and the estimated stellar oscillation noise^{9,15}, leading to a global error bar of 0.7–1.5 m s^{−1} per measurement. This does not include other noise sources that are intrinsic to the star, such as activity-related radial velocity jitter caused by cool spots at the stellar surface. However, HD 69830 exhibits low chromospheric activity, as can be seen by measuring the re-emission flux at the centre of the Ca II H and K lines. The normalized chromospheric emission index, $\log(R'_{\text{HK}})$ (ref. 16), obtained from our HARPS spectra, has an average value of −4.97, typical for old, quiet K dwarfs. Moreover, HD 69830 has a low projected rotational velocity: we measure $v \sin i = 1.1^{+0.5}_{-1.1}$ km s^{−1} using a calibration of the width of the cross-correlation function¹⁷ (where v is the stellar equatorial rotational velocity and i is the inclination angle of the stellar rotation axis relative to the line of sight). From these indicators we expect very low radial velocity jitter, probably below 1 m s^{−1}.

Orbital parameters for the three planets

The analysis of the radial velocity data reveals multi-periodic variations with a peak-to-peak amplitude of ~ 15 m s^{−1}. A close inspection of the radial velocity curve on short timescales clearly shows a sinusoidal modulation with a period of ~ 9 days, although successive maxima do not occur at the same radial velocity value, indicating that a second signal is present with a period of ~ 30 days. We first performed a two-keplerian fit to the data with starting values close to these two periods. The global r.m.s. dispersion of the residuals amounts to 1.57 m s^{−1} and the reduced χ^2 value is 4.19 (with 11 free parameters), meaning that this solution is not satisfactory. Moreover, the residuals around the fit are clearly not randomly distributed, but instead reveal another long-term periodicity at about 200 days. We therefore perform a three-keplerian fit to the radial velocities, which gives a much better result with a global (weighted) r.m.s. of 0.81 m s^{−1} and a reduced χ^2 value of 1.20 (16 free parameters). In this model, the inner planet has a period of 8.667 days, an eccentricity of 0.10 and a minimum mass of 10.2 $M_{\oplus}$ (where $M_{\oplus}$ is the Earth's mass). The second planet has a period of 31.56 days, an eccentricity of 0.13 and a minimum mass of 11.8 $M_{\oplus}$. Finally, the third planet has an orbital period of 197 days, an eccentricity of 0.07 and a minimum mass of 18.1 $M_{\oplus}$. The list of all orbital parameters for the system can be found in Table 1, while Fig. 1 shows the phase-folded radial velocity curves for the three planets.

Figure 2 shows two close-up views of the data and best-fit model as a function of time, together with the whole radial velocity curve after removal of the inner planets, thus revealing the long-term variations due to the third planet.

To check if our solution really gives the best fit to the data, we also explored the parameter space with a genetic algorithm. This technique, now routinely used to analyse radial velocity data, has the ability to find the absolute minimum on the χ^2 surface, avoiding the risk of getting trapped in a local minimum. In our case, the genetic algorithm yields orbital parameters that are indistinguishable from those we have found with the simple least-squares minimization, and confirms that the three-planet solution gives a superior fit compared to the two-planet model. Finally, we checked that the radial velocity variations are not partly due to stellar radial velocity jitter by computing the bisector velocity span of the cross-correlation function, which traces possible line shape variations¹⁸. The bisector turns out to be stable at the level of 0.81 m s^{−1} and is not correlated with any of the orbital periods, including the 31.6-day period which is close to the rotation period of the star (~ 35 days, estimated from the activity index¹⁶). If the signal was stellar in origin, bisector variations would occur with an amplitude similar to the radial velocity variations (~ 5 m s^{−1} for the second planet), and the bisector signal would vary in phase with the stellar rotation period. None of these signatures are observed in our case. We are thus confident that the radial velocity signal is indeed due to orbiting planetary companions.

Dynamical stability of the system

The multiple planetary system around HD 69830 is unique in that it is the first reported to be composed only of Neptune-mass objects, at least within a few astronomical units. Indeed, the time span of the observations (~ 800 days) and the precision of the measurements allow us to exclude any Saturn-mass planet orbiting within ~ 4 AU. When discovering a new system, the immediate question arising is whether or not it is dynamically stable over Myr to Gyr timescales. At first glance, the low planetary masses and small eccentricities suggest a high probability that the HD 69830 system is indeed stable. To investigate this point more thoroughly, we performed numerical N -body integrations¹⁹ assuming coplanarity of the orbits and two different inclination angles (corresponding to different true planetary masses). For both inclinations $i = 90^\circ$ (edge-on) and $i = 1^\circ$ (pole-on), the system turns out to remain stable over at least 1 Gyr. Whereas long-term stability could be expected in the minimum-mass case ($i = 90^\circ$, see Fig. 3a), it is notable that the system survives even if the true planetary masses lie in the Jupiter-mass range.

We also considered the possibility that an asteroid belt might exist within ~ 1 AU of the star, as suggested by the recent observations of the Spitzer Space Telescope⁸. Given the orbits of the three planets, the

Table 1 | Orbital and physical parameters of the planets in the HD 69830 system

Parameter	Planet		
	HD 69830 b	HD 69830 c	HD 69830 d
Orbital period (d)	8.667 $\pm$ 0.003	31.56 $\pm$ 0.04	197 $\pm$ 3
Time of periastron (BJD)	2453496.8 $\pm$ 0.6	2453469.6 $\pm$ 2.8	2453358 $\pm$ 34
Eccentricity	0.10 $\pm$ 0.04	0.13 $\pm$ 0.06	0.07 $\pm$ 0.07
Longitude of periastron (°)	340 $\pm$ 26	221 $\pm$ 35	224 $\pm$ 61
Velocity semi-amplitude (m s ^{−1})	3.51 $\pm$ 0.15	2.66 $\pm$ 0.16	2.20 $\pm$ 0.19
Semi-major axis (AU)	0.0785	0.186	0.630
Minimum mass ($M_{\oplus}$)	10.2	11.8	18.1
Number of data points	74		
O – C residuals (m s ^{−1})	0.81 (overall) / 1.50 (early) / 0.64 (late)		
Reduced χ^2 value	1.20		

The parameters and their formal 1 σ error bars are those given by the best three-keplerian fit to the data. We checked that planet-planet interactions do not significantly influence the orbital parameters by also performing N -body fits to the data with various inclination angles. As expected, dynamical interactions are so weak that they can be neglected over the short time span of the data, and we therefore use the three-keplerian fit in this paper. The r.m.s. of the observed minus calculated (O – C) residuals is given separately for early and late measurements to illustrate the higher quality of more recent data points. Note that the radial velocity semi-amplitudes of the three planets are the smallest measured to date, showing the potential of the radial velocity method to detect terrestrial planets close to their parent star. BJD, barycentric Julian date.

obvious question is whether or not there are zones of dynamical stability in the system where small bodies could survive in spite of the perturbing effects of the planets. For a uniform grid in semi-major axis (from 0.07 AU to 1.20 AU) and in eccentricity (from 0 to 0.9), massless particles were numerically integrated over two consecutive 1,000-year time intervals together with the three planets in the system. The variation of the mean motion frequency over the two time intervals provides a stability criterion for the particles²⁰. The results, shown in Fig. 3b, indicate that two regions seem to be stable enough to harbour an asteroid belt: an inner region between 0.3 AU and 0.5 AU, and the outer region beyond 0.8 AU (assuming there is no massive, as-yet undetected, planet further away). The observed high temperature of the emitting grains (~ 400 K) seems to favour the

hypothesis of a debris disk in the inner region, although the present observations may not allow us to clearly decide for one of the two regions. Conversely, the presence of a stable debris disk within ~ 1 AU of the star can be used to constrain the inclination of the system, as too-massive planets would have ejected all other bodies out of the central regions. Simulations show that no particle survives in the inner region for $i < 3^\circ$, corresponding to planetary masses in the Jupiter-mass range. Although not particularly strong, this constraint shows that dynamical studies are able to deliver valuable information for the characterization of planetary systems.

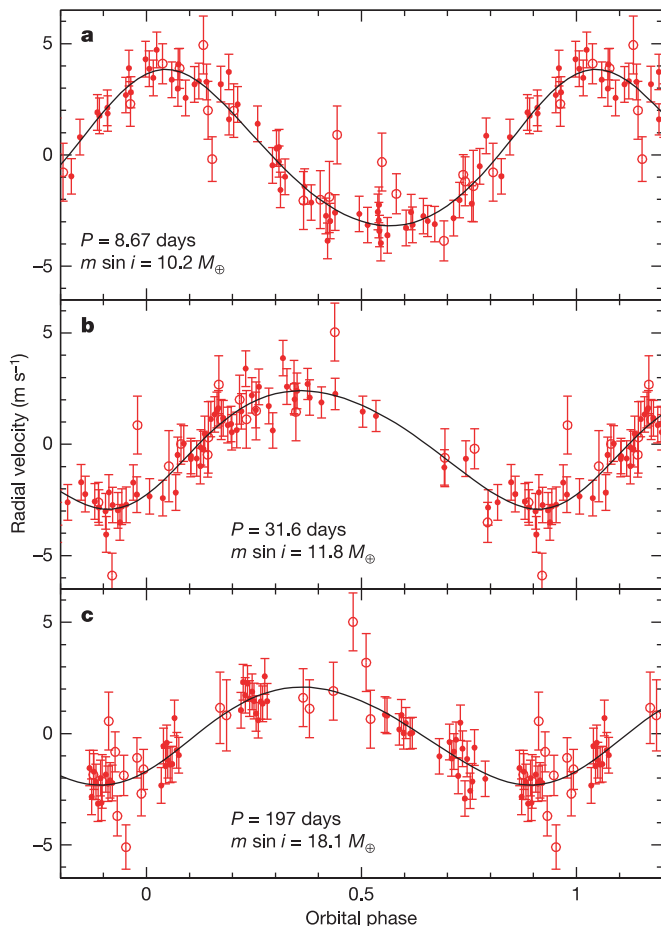


Figure 1 | Phase-folded radial velocity curves for the three planets. In each case, the contribution of the two other planets has been subtracted. The orbital periods, P , are 8.67, 31.6 and 197 days, for the inner (a), intermediate (b) and outer (c) planet, respectively. The radial velocity semi-amplitudes range from 3.5 to 2.2 m s^{-1} , corresponding to minimum masses $m \sin i$ of $10.2 M_{\oplus}$, $11.8 M_{\oplus}$ and $18.1 M_{\oplus}$ (here $M_{\oplus}$ is the Earth's mass, m is the actual planetary mass and i is the inclination angle of the system). The integration time was 4 min on average for the first 18 measurements (shown as open circles), and was increased to 15 min for the remaining points (filled circles). The latter measurements are of much higher quality for the following reasons: lower photon noise (from 0.4 to 0.2 m s^{-1}), improved guiding accuracy (from ~ 1.0 to 0.3 m s^{-1}), lower wavelength calibration error (from 0.8 to $\sim 0.3 \text{ m s}^{-1}$) and better averaging of the stellar p-mode oscillations (which have characteristic periods of a few minutes and individual amplitudes of a few tens of cm s^{-1} that may add up to a few m s^{-1})^{9,15}. For the K0 dwarf HD 69830, we estimate that the oscillation noise is between 0.2 and 0.8 m s^{-1} depending on the exposure time. Combining all these error sources in quadrature, we obtain final 1σ error bars between 0.7 and 1.5 m s^{-1} .

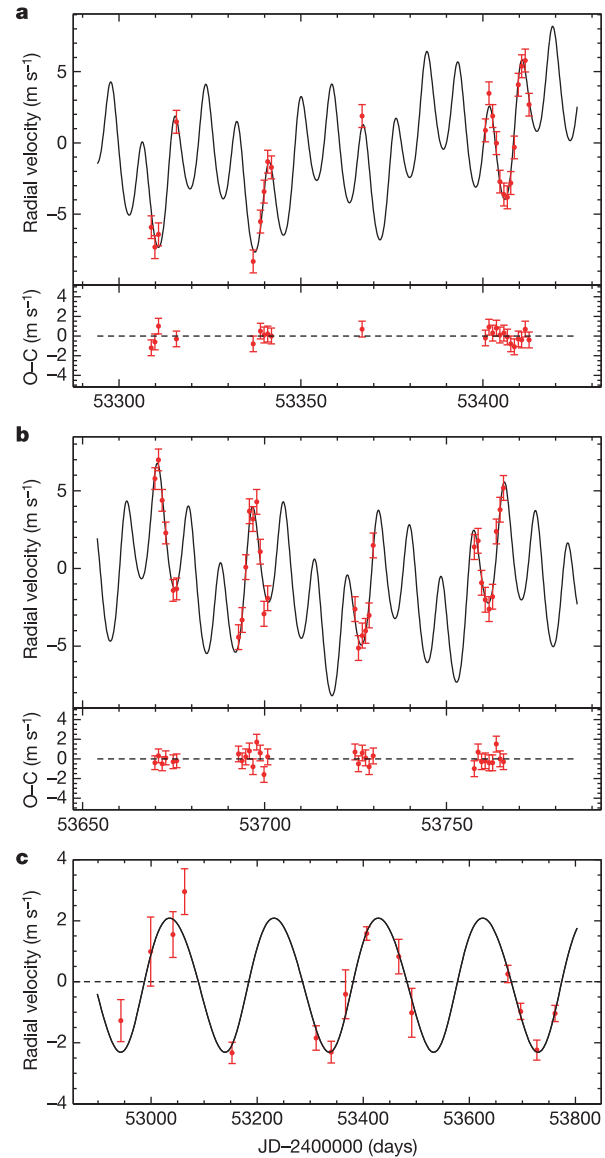


Figure 2 | Radial velocity curve as a function of time. a, b, Close-up views of the data, showing the cumulative signal of three planets. The short-period planet ($P = 8.67$ days) shows up as a high-frequency modulation, whereas the intermediate planet ($P = 31.6$ days) is revealed through the varying values of successive minima and maxima. The outer planet ($P = 197$ days) is not easily seen on these magnified views, but its presence becomes clear when removing the signal of the inner planets and binning the data points (one per observing run), as shown in c. Note that only high-quality radial velocity measurements are able to fully resolve this system. The weighted r.m.s. of the residuals around the best-fit model is 0.81 m s^{-1} and becomes as low as 0.64 m s^{-1} when considering only the more recent, higher-quality data points. O–C, observed minus calculated; JD, Julian date. Error bars are 1σ .

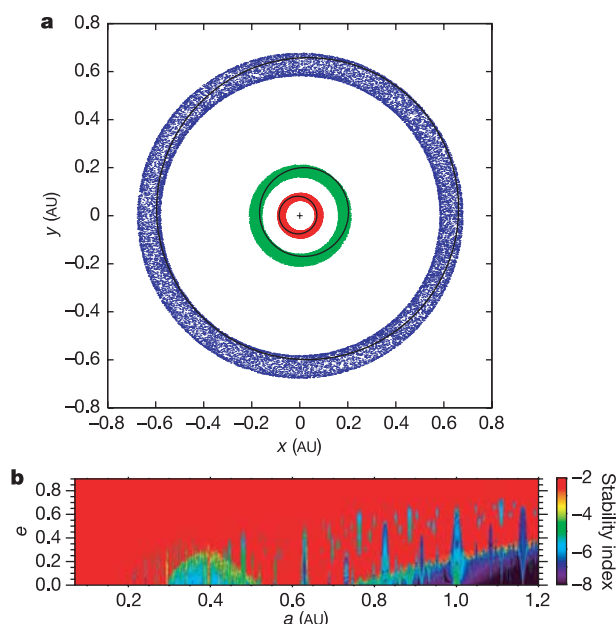


Figure 3 | Dynamical study of the HD 69830 system. **a**, Long-term evolution of the orbits of the three planets starting with the orbital solution from Table 1 and inclination equal to 90° . The panel shows a face-on view of the system; x and y are spatial coordinates in a frame centred on the star. Present orbital solutions are traced with solid lines and each dot corresponds to the position of the planet every 50,000 years. The system remained stable for at least 1 Gyr. The semi-major axes are constant, and the eccentricities undergo small variations ($0.05 < e_b < 0.20$, $0 < e_c < 0.14$ and $0.069 < e_d < 0.077$, where e_b , e_c and e_d denote the eccentricity of the inner, intermediate and outer planet, respectively). The fundamental periods related to the precession of the perihelion are respectively 5,266 yr, 15,855 yr and 148,000 yr. **b**, Stability analysis for massless particles. For a uniform grid of 0.001 AU in semi-major axis (a , from 0.07 to 1.20 AU), and 0.05 in eccentricity (e , from 0 to 0.9), massless particles are numerically integrated over two consecutive 1,000-year time intervals together with the three planets in the system. The variation of the mean motion frequency over the two time intervals provides a stability criterion for the particles²⁰. The colour grid corresponds to this stability index, red denoting the most unstable orbits with a close encounter with the central star or a planet, while dark blue corresponds to very stable orbits. The particles can survive for an extended time in a region between 0.3 and 0.5 AU, or in the more stable region beyond 0.8 AU. As for the Solar System main asteroid belt, several mean motion resonances with the outermost planet can be identified that would create gaps or accumulation in a potential asteroid belt: 1:2 (~ 0.40 AU), 2:3 (~ 0.48 AU), 1:1 (~ 0.63 AU), 3:2 (~ 0.82 AU) and 2:1 (~ 1.00 AU).

Formation and composition of the planets

The discovery of the HD 69830 system also represents a milestone in the understanding of planet formation. In the so-called core accretion model²¹, planetesimals accrete material from the protoplanetary disk, forming first a rocky or icy core with a mass up to $10\text{--}15 M_\oplus$. After that, a runaway gas accretion phase follows, which rapidly leads to the formation of a gaseous giant planet provided gas accretion starts before the evaporation of the disk. Simultaneously, inward migration due to interactions with the disk decreases the orbital semi-major axis. The final fate of the planet will be determined by the subtle balance between accretion, disk evaporation and migration timescales^{22,23}.

Using the models of ref. 22, we performed a large number of simulations aiming at reproducing the HD 69830 system. We assumed a stellar mass of $0.86 M_\odot$ and slightly sub-solar metallicity, but did not take into account the possible evaporation of the planets^{24,25}. We found that in order to reproduce the mass and semi-major axis of these planets, gas disk masses (between 0.07 AU and 30 AU) of $0.04\text{--}0.07 M_\odot$ and disk lifetimes of 1–3 Myr are required. These values are compatible with the observed ones^{26,27}.

We also found that to account for planets with masses between $10 M_\oplus$ and $20 M_\oplus$ at 0.2 AU, a significant amount of migration had to occur (at least one-tenth of the value analytically predicted for laminar disks²⁸). The starting positions of the embryos of the inner planets are found to be inside the ice line. Thus, these two bodies have grown by accreting essentially rocky planetesimals and some amount of gas. Obviously, if these planets lose mass owing to evaporation, a larger initial mass would be required to account for the present mass, implying that the embryos formed at larger distances. While the innermost planet would probably still start well within the ice line, the second one might accrete some limited amount of ices. The exact ratio between core mass and envelope mass depends on whether or not the disk of solids is depleted close to the central star (due to either the high temperatures in the inner disk, or to the formation process of the planets itself), as depletion in planetesimals triggers gas accretion²⁹. Assuming no depletion, the envelope of the inner planets is likely to remain very small, whereas strong depletion could lead to a substantial mass of accreted gas (up to $5 M_\oplus$).

The embryo of the outermost planet started beyond the ice line and therefore accreted a large amount of ices. Again, the core-to-envelope mass ratio depends on the possible depletion of the disk of solids by the formation of the second embryo. In order for the final mass to remain around $20 M_\oplus$, the Kelvin-Helmholtz time must be of the order of the disk lifetime (a few Myr), which limits the mass of the planet at this time to about $9\text{--}10 M_\oplus$ (ref. 23). Therefore, in the case of strong disk depletion by the second embryo, the final planet probably has a core-to-envelope mass ratio of about unity, while low depletion would lead to a smaller amount of accreted gas. Interestingly, this planet appears to be located near the inner edge of the habitable zone, where liquid water can exist at the surface of rocky/icy bodies³⁰. Indeed, the habitable zone is expected to be shifted closer to the star compared to our Solar System owing to the lower luminosity of HD 69830. Although this $\sim 20 M_\oplus$ planet is probably not telluric, its discovery opens the way to an exciting topic in astronomy: the characterization of low-mass planets in the habitable zone of solar-type stars.

Obviously, the planetary system around HD 69830 deserves thorough theoretical and observational investigations owing to its many interesting properties. It represents an important step forward in the characterization of planetary systems and will certainly help us to understand their huge diversity better.

Received 10 March; accepted 18 April 2006.

- Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
- Udry, S., Fischer, D. A. & Queloz, D. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) (Univ. Arizona Press, Tucson, in the press).
- Schneider, J. The Extrasolar Planets Encyclopaedia. (<http://exoplanet.eu>) (10 March 2006).
- Perryman, M. A. C. *et al.* The HIPPARCOS catalogue. *Astron. Astrophys.* **323**, L49–L52 (1997).
- Santos, N. C., Israelian, G. & Mayor, M. Spectroscopic [Fe/H] for 98 extra-solar planet-host stars. Exploring the probability of planet formation. *Astron. Astrophys.* **415**, 1153–1166 (2004).
- Schaller, G., Schaerer, D., Meynet, G. & Maeder, A. New grids of stellar models from 0.8 to 120 solar masses at $Z = 0.020$ and $Z = 0.001$. *Astron. Astrophys. Suppl. Ser.* **96**, 269–331 (1992).
- Girardi, L., Bressan, A., Bertelli, G. & Chiosi, C. Evolutionary tracks and isochrones for low- and intermediate-mass stars: From 0.15 to $7 M_\odot$, and from $Z = 0.0004$ to 0.03. *Astron. Astrophys. Suppl. Ser.* **141**, 371–383 (2000).
- Beichman, C. A. *et al.* An excess due to small grains around the nearby K0 V star HD 69830: Asteroid or cometary debris? *Astrophys. J.* **626**, 1061–1069 (2005).
- Mayor, M. *et al.* Setting new standards with HARPS. *Messenger* **114**, 20–24 (2003).
- Pepe, F. *et al.* The HARPS search for southern extra-solar planets. I. HD 330075 b: A new “hot Jupiter”. *Astron. Astrophys.* **423**, 385–389 (2004).
- Santos, N. C. *et al.* The HARPS survey for southern extra-solar planets. II. A 14 Earth-masses exoplanet around μ Arae. *Astron. Astrophys.* **426**, L19–L23 (2004).
- Lovis, C. *et al.* The HARPS search for southern extra-solar planets. III. Three Saturn-mass planets around HD 93083, HD 101930 and HD 102117. *Astron. Astrophys.* **437**, 1121–1126 (2005).

13. Udry, S. *et al.* The HARPS search for southern extra-solar planets. V. A 14 Earth-masses planet orbiting HD 4308. *Astron. Astrophys.* **447**, 361–367 (2006).
14. Baranne, A. *et al.* ELODIE: A spectrograph for accurate radial velocity measurements. *Astron. Astrophys.* **119** (Suppl.), 373–390 (1996).
15. Bouchy, F., Bazot, M., Santos, N. C., Vauclair, S. & Sosnowska, D. Asteroseismology of the planet-hosting star μ Arae. I. The acoustic spectrum. *Astron. Astrophys.* **440**, 609–614 (2005).
16. Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K. & Vaughan, A. H. Rotation, convection, and magnetic activity in lower main-sequence stars. *Astrophys. J.* **279**, 763–777 (1984).
17. Santos, N. C. *et al.* The CORALIE survey for southern extra-solar planets. IX. A 1.3-day period brown dwarf disguised as a planet. *Astron. Astrophys.* **392**, 215–229 (2002).
18. Queloz, D. *et al.* No planet for HD 166435. *Astron. Astrophys.* **379**, 279–287 (2001).
19. Correia, A. C. M. *et al.* The CORALIE survey for southern extra-solar planets. XIII. A pair of planets around HD 202206 or a circumbinary planet? *Astron. Astrophys.* **440**, 751–758 (2005).
20. Laskar, J. Frequency analysis for multidimensional systems—global dynamics and diffusion. *Physica D* **67**, 257–281 (1993).
21. Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
22. Alibert, Y., Mordasini, C., Benz, W. & Winisdoerffer, C. Models of giant planet formation with migration and disc evolution. *Astron. Astrophys.* **434**, 343–353 (2005).
23. Ida, S. & Lin, D. N. C. Toward a deterministic model of planetary formation. I. A desert in the mass and semimajor axis distributions of extrasolar planets. *Astrophys. J.* **604**, 388–413 (2004).
24. Lecavelier des Etangs, A., Vidal-Madjar, A., McConnell, J. C. & Hébrard, G. Atmospheric escape from hot Jupiters. *Astron. Astrophys.* **418**, L1–L4 (2004).
25. Baraffe, I., Alibert, Y., Chabrier, G. & Benz, W. Birth and fate of hot-Neptune planets. *Astron. Astrophys.* (in the press); preprint at (<http://arXiv.org/astro-ph/0512091>) (2005).
26. Haisch, K. E., Lada, E. A. & Lada, C. J. Disk frequencies and lifetimes in young clusters. *Astrophys. J.* **553**, L153–L156 (2001).
27. Beckwith, S. V. W. & Sargent, A. I. Circumstellar disks and the search for neighbouring planetary systems. *Nature* **383**, 139–144 (1996).
28. Tanaka, H., Takeuchi, T. & Ward, W. R. Three-dimensional interaction between a planet and an isothermal gaseous disk. I. Corotation and Lindblad torques and planet migration. *Astrophys. J.* **565**, 1257–1274 (2002).
29. Alibert, Y., Mousis, O., Mordasini, C. & Benz, W. New Jupiter and Saturn formation models meet observations. *Astrophys. J.* **626**, L57–L60 (2005).
30. Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. Habitable zones around main sequence stars. *Icarus* **101**, 108–128 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The data presented here were obtained with the ESO 3.6-m telescope at La Silla Paranal Observatory, Chile. We thank O. Tamuz for the development of the genetic algorithm code, and J. Couetdic for his help in studying the stability of the asteroid belt. We thank the Swiss National Science Foundation (FNRS), the Centre National de la Recherche Scientifique (France) and the Fundação para Ciência e a Tecnologia (Portugal) for their continuous support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.L. (christophe.lovis@obs.unige.ch) or M.M. (michel.mayor@obs.unige.ch).

ARTICLES

Evolution of an obligate social cheater to a superior cooperator

Francesca Fiegna¹, Yuen-Tsu N. Yu¹, Supriya V. Kadam¹ & Gregory J. Velicer¹

Obligate relationships have evolved many times and can be parasitic or mutualistic. Obligate organisms rely on others to survive and thus coevolve with their host or partner. An important but little explored question is whether obligate status is an evolutionarily terminal condition or whether obligate lineages can evolve back to an autonomous lifestyle. The bacterium *Myxococcus xanthus* survives starvation by the social development of spore-bearing fruiting bodies. Some *M. xanthus* genotypes defective at fruiting body development in isolation can nonetheless exploit proficient genotypes in chimaeric groups. Here we report an evolutionary transition from obligate dependence on an altruistic host to an autonomous mode of social cooperation. This restoration of social independence was caused by a single mutation of large effect that confers fitness superiority over both ancestral genotypes, including immunity from exploitation by the ancestral cheater. Thus, a temporary state of obligate cheating served as an evolutionary stepping-stone to a novel state of autonomous social dominance.

Obligate relationships are common in biology. Higher eukaryotes host a vast array of obligate parasites and mutualists that include both microbes and larger organisms^{1–6}. For example, numerous bacteria are obligate pathogens or symbionts^{7,8}, many bird species are obligate brood parasites⁹ and a variety of insects are social parasites that rely on workers of another species to raise their offspring^{10–13}. These radically distinct obligate systems share the common theme of dependence on a host or partner (which might be of the same or of a different species) for evolutionary success despite their divergent mechanisms and strategies.

At least in microbes, extensive gene inactivation or loss often occurs during long-term evolution in an obligate state of mutualism or parasitism^{7,8}, making reversion to free-living status highly unlikely after extended periods of evolution. Although these and other long-term patterns of coevolution in obligate systems have been detected^{9,14,15}, little is known about the initial stages of obligation in any system¹⁶. It remains unclear how frequently genetic transitions to obligate status are contingently irreversible¹⁷ or rather might reverse during early stages of obligate coevolution, either directly to the exact ancestral genotype or by a novel mutational pathway^{18,19}. In particular, little is known about the potential of obligate social cheaters to evolve new forms of social cooperation that do not depend on the prior social host.

Myxococcus xanthus is the best characterized species of the soil-dwelling myxobacteria, which are distinguished by cooperative predation on other microbes and social development into a wide variety of fruiting body morphologies²⁰. Predation is accomplished by swarming packs of cells that secrete toxic and lytic metabolites that kill and degrade prey organisms and thereby generate an extracellular, public pool of resources²¹. Both predation and development are facilitated by two motility systems, one of which is social in nature and employs type-IV pili that mediate cell–cell interactions²². Upon amino-acid deprivation, *M. xanthus* cells aggregate into local groups of ~100,000 individuals that exchange intercellular signals to construct spore-bearing fruiting bodies²³. Importantly, only a minority of cells survive development after differentiation into stress-resistant spores²⁴, and thus proficiency at competition for limited sporulation

‘slots’ can be a major fitness component for distinct genotypes undergoing co-development in chimaeric fruiting bodies^{25,26}. This microbial system (as well as others^{27,28}) readily allows real-time observation of evolutionary changes in social traits^{29,30}.

A variety of socially exploitative relationships among microbes have been described^{25,26,31–34}, including developmentally-defective genotypes of *M. xanthus* that are capable of cheating on their developmentally-proficient progenitor during starvation^{26,35}. Several such cheaters were isolated from *M. xanthus* populations that had undergone 1,000 generations of evolution in nutrient-rich liquid medium in which positive selection for proficient motility or development was absent^{26,30}. Under these conditions all populations incurred major evolutionary losses in social ability, with most lineages becoming defective in both social motility and developmental sporulation.

In mixtures with the developmentally proficient ancestor, however, some of these developmentally incompetent genotypes are able to cheat on their social benefactor by sporulating even more efficiently than the ancestral cooperator^{26,35}. One such cheater genotype, here designated ‘OC’ (Obligate Cheater), fails to produce viable spores in clonal isolation³⁵. In this experimental system, OC is thus an obligate social cheater that requires the presence of a developmentally proficient social host (for example, strain GJV2, a marked variant of OC’s ancestor GJV1) to avoid extinction during starvation. GJV2 is altruistic in an evolutionary sense because it benefits the fitness of OC to its own competitive disadvantage.

Evolutionary restoration of social autonomy

This obligate cheater (OC) was allowed to compete against the socially competent strain GJV2 over six sequential competition cycles, with each cycle consisting of two phases—development on starvation agar followed immediately by growth in liquid medium³⁶. Development and growth phase cycles are hereafter referred to as D1–D6 and G1–G6, respectively. OC began as 1% of the initial mixed population, and the two competitors were marked with resistance to distinct antibiotics (kanamycin for OC and rifampicin for GJV2), which allowed tracking of genotype frequencies throughout the

¹Max Planck Institute for Developmental Biology, Spemannstrasse 35, D-72076 Tübingen, Germany.

competition. In one competition that we focus on here, descendants of OC re-evolved the ability to sporulate independently of GJV2 and overtook the population (Fig. 1).

As a minority in mixed culture with GJV2, the initial OC cheater genotype sporulates more efficiently than GJV2 during development and also shows an advantage during growth in liquid medium^{35,36}. Combined, these competitive advantages allowed the OC population to rise rapidly from an initial frequency of 0.01 to majority status (Fig. 1). During the second and third rounds of development (D2 and D3), the increasing frequency of developmentally-defective OC cells led to large reductions in total spore production (of both competitors combined) relative to the first round (D1) in which GJV2 began at a frequency of 0.99. An even more severe sporulation bottleneck occurred during D4, which fewer than 20 individuals (the lower limit of detection) appear to have survived. OC descendants overtook the population after the D4 bottleneck.

If the OC-derived population entering D5 had undergone no evolutionary change since the onset of competition, it should have produced zero or extremely few spores during D5 owing to the developmental defect of OC and the absence of sufficient GJV2 cells necessary to rescue OC sporulation. Surprisingly, the previously defective OC-derived population produced spores at a very high level during D5 and D6, despite the absence of GJV2 cells (Fig. 1). Total spore production during D5 and D6 was 2.5- and 2.8-fold higher, respectively, than during D1 when GJV2 was present as 99% of the population ($P < 0.05$ in both cases, t -test).

The radical change in developmental phenotype of the kanamycin-resistant OC-derived population suggested that a spontaneous mutant of OC that had regained sporulation proficiency originated before D4 and survived the D4 bottleneck at high frequency. Nine kanamycin-resistant, rifampicin-sensitive clones (OC-derived) were isolated from culture samples frozen at the end of each growth phase throughout the competition experiment. All OC-derived clones from population samples before D4 (G1–G3) failed to aggregate or produce spores in clonal populations (data not shown). All clones

isolated after the D4 crash (G4–G6) showed the positive developmental phenotype exhibited by the original competition population in D5 and D6 (Fig. 1). Hereafter, we refer to the sporulation-proficient genotype derived from OC that arose from the D4 population crash as 'PX' (Phoenix). One PX clone from the G4 population was selected for further analysis. In clonal cultures, PX sporulated sevenfold more efficiently than GJV2 (Fig. 2, $P < 0.01$, t -test). We estimate that ≤ 60 generations of growth occurred between the initial experimental inoculation of OC and the detection of PX.

Usurper dominance

Because OC dominated the mixed OC-derived population at the onset of D4 (frequency > 0.89), the very large gain in relative frequency by PX during the fourth cycle of development and growth suggested that PX has a large survival advantage over OC during development when PX is rare and OC is common. This is in fact the case, as the sporulation efficiency of PX is $> 1,000$ -fold higher than that of OC in direct competition between these two genotypes when PX is 10% of the initial population (Fig. 3).

The novel PX genotype appears to have fully displaced its immediate ancestor OC as well as the GJV2 population, suggesting that PX is a superior competitor to both. To examine its fitness against these genotypes more thoroughly, PX was mixed at multiple initial frequencies with both competitors independently and its relative sporulation efficiency estimated after one round of development (Fig. 3). PX was developmentally superior to both competitors at all frequencies from 0.1 to 0.9, but showed opposite relationships between frequency and degree of superiority across the two competitors. The advantage of PX over GJV2 increased ~ 400 -fold as PX frequency increased from 0.1 to 0.9, whereas the degree of its advantage over OC decreased ~ 700 -fold over the same frequency range.

The frequency-dependent relationships between OC and the two proficient strains GJV2 and PX were qualitatively similar, with total social productivity (that is, the combined spore production of both competitors in each pairing) and OC fitness (Fig. 3) both being inversely correlated with OC frequency (total sporulation and OC versus GJV2 data not shown). One significant exception is that even at low OC frequencies, extracellular complementation of OC sporulation by PX was only partial (that is, OC gains no cheating advantage against PX from its defect), whereas GJV2 supports a true cheating phenotype in which defection by OC translates into a fitness advantage. Thus, the mutational pathway that restores development in PX also renders PX competitively immune from exploitation by its cheating parent. Such immunity from cheating—or some other form of cheater control^{17,37–40}—is necessary for any novel cooperative strategy to succeed when defectors are present.

The result that PX superiority over GJV2 increases dramatically as a positive function of PX frequency suggests that the composition or amount of extracellular compounds that affect development produced by the two genotypes is asymmetric, and that the sporulation efficiency of each strain is greatest when its own molecular profile dominates the extracellular social matrix. Because PX is superior at all frequencies examined, it should rise to dominance in extended

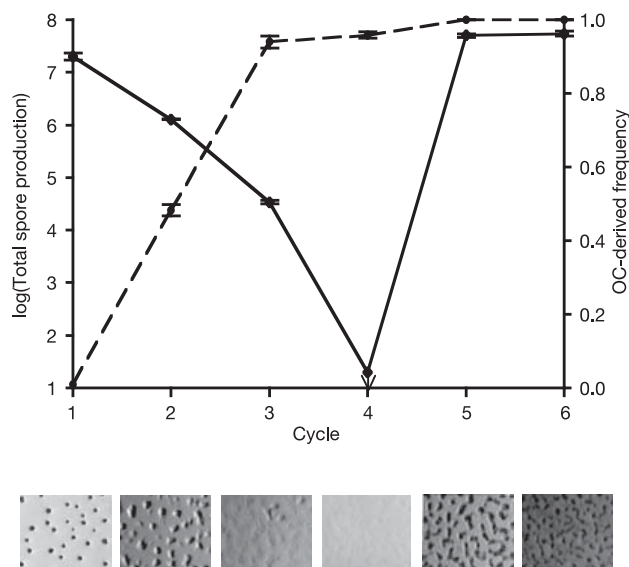


Figure 1 | Total spore production, OC-derived sub-population frequencies and developmental population phenotypes during the original six-cycle competition between OC and GJV2. OC-derived frequencies (dashed line, right axis) represent sub-population frequencies at the beginning of each respective developmental cycle (x axis), whereas spore production values (solid line, left axis) represent total spore population sizes at the end of each developmental stage. Developmental phenotypes (images) are shown in sequential order (cycles 1 to 6, left to right). Error bars indicate 95% confidence intervals based on two independent samples of the population at each time point. The arrow indicates that no spores were present at the lower limit of detection in cycle 4.

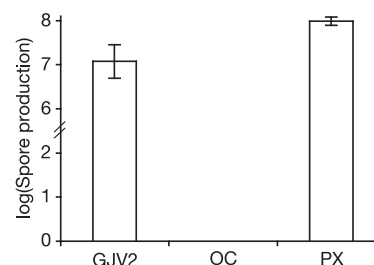


Figure 2 | Pure-culture spore production of strains GJV2, OC and PX. No spores were produced by strain OC. Error bars indicate 95% confidence intervals.

competitions against either competitor over sequential cycles of development. In a subsequent competition with GJV2, PX rapidly rose from 1% of the initial population to near fixation over six rounds of development (Fig. 4).

The developmental superiority of PX over GJV2 does not derive from an ability to sporulate more efficiently in the presence of GJV2 than in isolation (that is, it does not exploit GJV2, data not shown). Rather, its dominance is due exclusively to superior independent sporulation as well as its ability, at high frequencies, to decrease the efficiency of GJV2 sporulation. At low frequencies, OC is also superior to GJV2, but its superiority is due exclusively to mixing-specific exploitation of GJV2. Thus the evolutionary transition from OC to PX was a restoration of social independence that resulted in a loss of ability to exploit GJV2. Such exploitation, however, is not necessary for PX to be socially dominant.

One mutation restores social development

The rapidity of the evolutionary transition from OC to PX suggested that a single mutation of large phenotypic effect may have caused the observed adaptive jump in social and developmental complexity⁴¹. To identify mutations potentially responsible for restoring social independence in PX, the PX genome was sequenced to ~19-fold coverage⁴². Only a single mutation was found that distinguishes the PX genome from that of its immediate OC ancestor (out of 15 total evolutionary mutational differences between PX and GJV1, the parent of GJV2 and direct ancestor of OC)⁴². This PX-specific mutation altered the central position of a seven-base cytosine run to adenosine, and is located 128 bases above the annotated start codon of a predicted GNAT-family acetyltransferase⁴³, one of more than 30 in the *M. xanthus* genome (W. C. Nierman *et al.*, unpublished work). The roles of this and other GNAT-acetyltransferases in *M. xanthus* physiology and development have not been characterized. The mutation is absent from all 27 OC-derived clones isolated before the D4 crash that fail to sporulate as clonal cultures. In contrast, all 27 clones isolated from after the D4 crash that exhibit the PX phenotype were found to bear this mutation, which strongly suggested a causal relationship between the mutation and the PX phenotype.

Nonetheless, the PX phenotype might have been conferred by reversion of a mutation present in OC back to its ancestral state in GJV1. Such mutations would not have been detected by sequencing the PX genome, which was compared to the ancestral genome template⁴². To directly test the candidate mutation's function, a

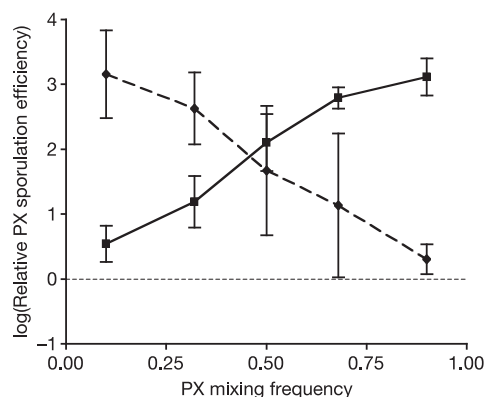


Figure 3 | Log-transformed ratios of PX sporulation efficiency relative to that of GJV2 and OC in direct pairwise competitions at multiple mixing frequencies. The sporulation efficiency of a strain is the frequency of its population at the onset of starvation that survives development as stress-resistant spores. PX sporulates more efficiently than both competitors (GJV2, solid line, and OC, dashed line) at all mixing frequencies (that is, the log-transformed values of PX relative efficiencies are positive in all cases). The horizontal short-dashed line indicates the value (zero) at which two competitors would have equal developmental fitness. Error bars indicate 95% confidence intervals.

markerless transfer of the mutation from PX to GVB207.3 (the immediate unmarked parent of strain OC, see Methods) was performed. A counter-selectable marker system was employed^{29,44} in which a plasmid bearing the PX mutation and surrounding sequence was integrated into and then excised from the GVB207.3 genome, thus resulting in clones with either the exact PX sequence (GVB207.3/PX+) or the ancestral sequence (GVB207.3/PX-). Three independently isolated GVB207.3/PX+ clones all showed the PX fruiting and sporulation phenotype (Fig. 5). In pure culture assays comparing spore production by PX, the GVB207.3/PX+ clones and GJV1, the PX and GVB207.3/PX+ genotypes were both superior to GJV1 (8- and 12-fold, respectively, $P < 0.01$ in both cases, *t*-test), but were not significantly different from one another ($P = 0.19$). Three GVB207.3/PX- excisants retained the ancestral base at the mutation site and produced no fruiting bodies or spores (Fig. 5).

The PX mutation may confer its phenotype through *cis* regulation of the developmental expression of the acetyltransferase gene immediately below the mutation. During vegetative growth (and hence at the onset of starvation), expression of this gene was fivefold higher in PX than in a kanamycin-resistant sister-variant of strain GJV2 ('GJV10') and threefold higher than in OC (Fig. 6). Expression remained significantly higher in PX than in GJV10 after six (fourfold difference) and twelve (twofold difference) hours of starvation ($P = 0.033$, 0.0072 and 0.0068 for comparisons at 0, 6 and 12 hours, respectively, *t*-tests on log₁₀-transformed data) (Fig. 6). Expression in PX increased slightly over the first six hours of development but decreased thereafter to only 65% of its initial level after 24 hours. In contrast, expression in the developmentally proficient ancestral variant increased steadily over 24 hours, after which it slightly exceeded expression in PX. Interestingly, expression of the acetyltransferase in OC is not defective relative to GJV10, but rather was significantly higher (1.7-fold difference, $P = 0.0035$) at the onset of starvation, after which it increased gradually to levels similar to the other two strains at 24 hours. If the increased acetyltransferase expression in PX is the cause of developmental restoration, it represents a novel transcriptional pathway by which proficient development can be accomplished.

Increased developmental expression of this acetyltransferase may cause developmental restoration in PX by modifying the expression of many other developmentally-regulated genes via an uncharacterized genetic pathway. Preliminary quantitative real-time PCR (polymerase chain reaction) studies of several genes with development-specific expression patterns show that expression of these genes during development changed greatly in the transition from OC to PX, indicating that the transition mutation has many and diverse regulatory effects (data not shown). Moreover, expression patterns of these genes often appear to differ significantly between GJV10 and PX, suggesting that PX development is associated with unique developmental transcription patterns that may be functionally significant.

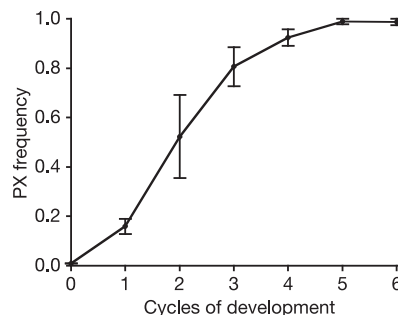


Figure 4 | Frequency of PX in direct competition with GJV2 over six cycles of development. The initial PX frequency was 0.01 and subsequent estimates were made after harvesting spores at the end of each round of development. Error bars indicate 95% confidence intervals.

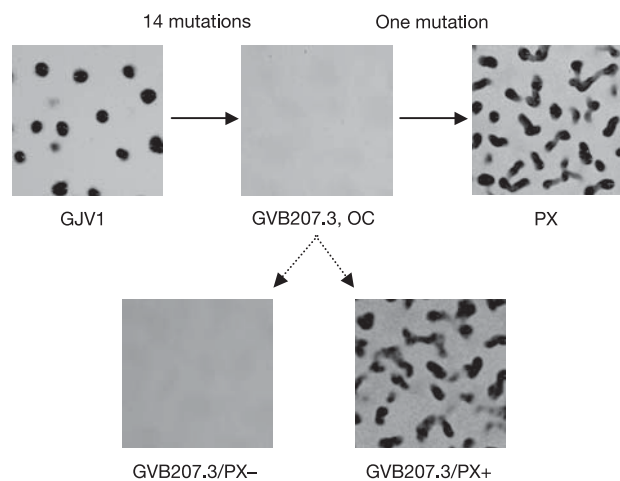


Figure 5 | Fruiting morphology effects of the PX-specific mutation. Strain GVB207.3 accumulated 14 mutational differences from its ancestor GJV1 over 1,000 generations of evolution, whereas strain PX was generated by only a single mutation during ~60 generations of evolution from strain OC (a marked derivative of GVB207.3). Transfer of the PX-specific mutation into the GVB207.3 genomic background fully confers the proficient PX fruiting and sporulation phenotypes (lower right, GVB207.3/PX+), whereas control strains lacking the mutation retain the defective GVB207.3 and OC phenotype (lower left, GVB207.3/PX-).

The discovery of this regulatory mutation reveals a genetic pathway allowing multicellular development that had not been identified by traditional screens for mutations that suppress developmental defects. This outcome highlights the power of experimental evolution, combined with effective mutation-identification methods, to not only illuminate evolutionary dynamics at both population and molecular levels, but also to generate novel insights into the basic biology of model organisms. It also highlights the variety of pathways by which cooperative behaviours might evolve. Development in PX was not restored by direct reversion of a loss-of-function mutation, but rather by a novel mutational pathway. This evolutionary transition has some similarities to a previously reported restoration of a defective social phenotype (an engineered defect in *M. xanthus* social motility) that was accomplished by compensatory mutation(s) that generated a novel mechanistic basis for an adaptive cooperative trait²⁹.

The importance of genomic background for expression of the PX phenotype is unclear. Ongoing studies seek to identify which of the 14 mutations that occurred during the 1,000 generations of evolution from GJV1 to GVB207.3 are responsible for developmental incompetence and cheating ability in OC, and whether the PX mutation requires epistatic interaction with any of these mutations to confer the PX phenotype. It will be of further interest to determine whether the PX mutation restores development only in genotypes with the same type of developmental defect as that of OC or is also able to confer developmental proficiency to a broader range of defective genotypes.

The likelihood of evolving out of an obligate relationship should decrease as a function of time spent in the obligate state. Gene functions provided by the sponsoring host or partner should be lost over time in the dependent lineage via selection or drift, making reversion to functional independence increasingly difficult. Before reaching an evolutionary point-of-no-return during obligate dependence, however, some mutations may generate novel independent forms that dominate over competing genotypes. Our results suggest that some extant lineages of free-living organisms may have emerged from temporary evolutionary periods in an obligate state by simple mutational pathways. The relative roles of mutations with small versus large effects on fitness and phenotype in determining evolutionary outcomes have long been investigated⁴⁵. The rapid transition

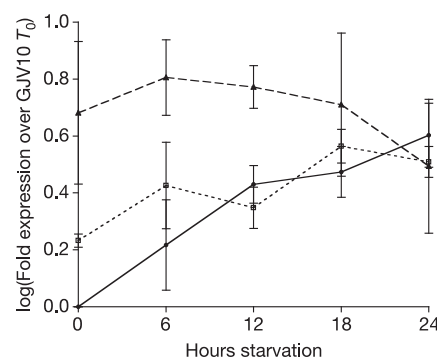


Figure 6 | Developmental expression of the putative acetyltransferase immediately downstream of the PX mutation. Expression levels are log-transformed estimates of fold-differences between GJV10 (a marked variant of GJV1, solid line), OC (dotted line) and PX (dashed line) at various starvation time-points relative to the expression level of GJV10 at the onset of starvation (T_0). Error bars indicate 95% confidence intervals.

from obligate cheating to autonomous social dominance reported here adds to our understanding of possible tempos and modes of adaptive evolution in social and developmental systems.

METHODS

Strain GJV2 is a spontaneous rifampicin-resistant mutant of GJV1, which is a clone of the standard laboratory strain DK1622 (ref. 46). (GVJ2 corresponds to 'R' and 'WT3' in refs 30 and 36, respectively.) Strain GVB207.3 ('S2' in ref. 35) is a 1,000-generation descendant of its ancestor GJV1 (refs 30, 35). Strain OC in this study (also 'S2/pDW79-' in ref. 35) is a kanamycin-resistant derivative of GVB207.3 generated by chromosomal integration of the plasmid pDW79 (refs 35, 47). The ability of OC to cheat is apparently not caused by pDW79, because a rifampicin-resistant mutant of GVB207.3 (with no plasmid) also exhibits cheating (data not shown).

The original competition in which PX evolved from OC was conducted as described in ref. 36, as was the subsequent competition between PX and GJV1 (Fig. 4) in which PX began as 1% of the population in three independent replicates. Briefly, mixed populations of 5×10^8 cells underwent five days of starvation on buffered agar and were then harvested and heated to kill non-spores. Half of the remaining spore population underwent germination and growth in liquid medium, whereas the other half was sonicated and diluted into selective agar to obtain spore counts of each competitor. This protocol was repeated for multi-cycle competitions. The single-cycle competitions between PX and GJV1 and between PX and a spontaneous rifampicin-resistant mutant of OC (Fig. 3) were performed in three replicate blocks. The presence or absence of the candidate mutation for the transition from OC to PX was examined in 54 OC-derived clones by standard PCR amplification and sequencing of the surrounding region.

Markerless allele exchange. Strain PX was sequenced to ~19-fold average coverage by 454 Life Sciences⁴² and reads were assembled against the complete 9,139,763 base pair (bp) genome sequence of strain DK1622 (GenBank accession CP000113). The sole discovered mutation unique to PX is at position 1,258,238 and is 32 bp upstream from a predicted GNAT-family acetyltransferase. A PCR fragment of the PX genome extending 452 bp above and 657 bp below the PX mutation was ligated into the cloning vector pCR-Blunt (Invitrogen), which bears a kanamycin-resistance gene, to generate pNY-PX.1. A *Bam*HI-*Eco*RV restriction fragment of pNY-PX.1 containing the cloned region was ligated into the plasmid pBJ113 (ref. 44) digested with *Bam*HI and *Hinc*II to create pNY-PX.2. Plasmid pBJ113 carries kanamycin resistance and *galK* genes that allow selection for plasmid integration and screening for vector excision, respectively, resulting in allele exchange⁴⁴. Strain GVB207.3 was electroporated with pNY-PX.2 and transformants with the integrated plasmid were selected on kanamycin agar and several transformants were plated onto 1% galactose-CTT agar. With the *galK* gene as a counter-selectable marker, galactose-resistant/kanamycin-sensitive colonies were obtained and screened for the PX mutation replacement in GVB207.3 by PCR and sequencing. Plasmid excision resulted in two markerless excisants types that either returned to the original GVB207.3 sequence (GVB207.3/PX-), or have the exchanged PX mutation (GVB207.3/PX+). The developmental morphology and sporulation levels of these excisants were compared to GJV1, GVB207.3, OC and PX as described above.

Real-time PCR. Developmental gene expression of the acetyltransferase downstream of the PX mutation was compared across PX, OC and GJV10, a derivative

of GJV1 marked with kanamycin resistance by transformation with the same plasmid (pDW79) used to generate strain OC from GVB207.3. Development was initiated as in the population assay, except that 16 100- μ l aliquots of resuspended culture per strain were deposited on buffered-agar plates for each time point harvest. Samples were harvested at 6, 12, 18 and 24 h in 1.5 ml TPM liquid⁴⁸ and transferred into 6 ml of Bacteria Protect Reagent (BPR, Qiagen). Fruiting bodies were dispersed, incubated at room temperature for 10 min, centrifuged at 4,000 g for 10 min (4 °C) and frozen after supernatant removal. For the T_0 (immediately prior to starvation) sample, a centrifuged sample of the original liquid culture used to initiate development was resuspended in BPR and processed in the same manner.

To isolate RNA, frozen pellets were thawed and resuspended in 4 ml of 30 mg ml⁻¹ lysozyme (Sigma) prepared in TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0) for cell lysis. RNA was extracted with a Qiagen Midi Kit. To remove DNA, RNA eluted in 250 μ l of RNase free water was supplemented with 25 μ l DNase buffer and 12.5 μ l of DNase I (Ambion), incubated at 37 °C for one hour, and cleaned using a Qiagen clean-up kit. Quality was checked by reverse transcriptase-PCR using the forward primer CTCCTCACCCATAAAATC ACCC and reverse primer CCTCGAAGGCCTCTGGA.

Two μ g of RNA treated with DNase I was used to synthesize random complementary DNAs using a reverse transcriptase kit (Applied Biosystems). Subsequently, real-time PCR reactions were prepared by adding 23 μ l of premix in the wells of a 96-well optical PCR plate (Applied Biosystems) and then adding 2 μ l of cDNA along with primer at a final concentration of 100 nM of gene-specific forward and reverse primers and 1x SYBR Green Mastermix (ABI). 25 μ l reaction mixes were placed in an Opticon Cycler (MJ Research), incubated at 95 °C for 10 min and subjected to 40 cycles of 15 s at 95 °C and 60 s at 60 °C, followed by one cycle of 15 s at 95 °C, 30 s at 60 °C and 15 s at 95 °C. The experiment was carried out in three independent biological blocks and two temporally independent technical replicates per block.

Received 3 January; accepted 23 February 2006.

- Pierce, N. E. *et al.* The ecology and evolution of ant association in the Lycaenidae (Lepidoptera). *Annu. Rev. Entomol.* **47**, 733–771 (2002).
- Keeling, P. J. & Fast, N. M. Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annu. Rev. Microbiol.* **56**, 93–116 (2002).
- Alexander, J., Satoskar, A. R. & Russell, D. G. Leishmania species: models of intracellular parasitism. *J. Cell Sci.* **112**, 2993–3002 (1999).
- Williamson, V. M. & Gleason, C. A. Plant-nematode interactions. *Curr. Opin. Plant Biol.* **6**, 327–333 (2003).
- Pellmyr, O. & Krenn, H. W. Origin of a complex key innovation in an obligate insect-plant mutualism. *Proc. Natl Acad. Sci. USA* **99**, 5498–5502 (2002).
- Lopez-Vaamonde, C., Rasplus, J. Y., Weiblen, G. D. & Cook, J. M. Molecular phylogenies of fig wasps: partial cocladogenesis of pollinators and parasites. *Mol. Phylogenet. Evol.* **21**, 55–71 (2001).
- Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583–586 (2002).
- Moran, N. A. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* **6**, 512–518 (2003).
- Rothstein, S. I. & Robinson, S. K. *Parasitic Birds and their Hosts: Studies in Coevolution* (Oxford Univ. Press, New York, 1998).
- Lorenzi, M. C., Cervo, R., Zacchi, F., Turillazzi, S. & Bagnères, A. G. Dynamics of chemical mimicry in the social parasite wasp *Polistes semenowi* (Hymenoptera: Vespidae). *Parasitology* **129**, 643–651 (2004).
- Sumner, S., Aanen, D. K., Delabie, J. & Boomsma, J. J. The evolution of social parasitism in Acromyrmex leaf-cutting ants: a test of Emery's rule. *Insectes Soc.* **51**, 37–42 (2004).
- Cervo, R., Macinai, V., Dechigi, F. & Turillazzi, S. Fast growth of immature brood in a social parasite wasp: A convergent evolution between avian and insect cuckoos. *Am. Nat.* **164**, 814–820 (2004).
- Foitzik, S., Fischer, B. & Heinze, J. Arms races between social parasites and their hosts: geographic patterns of manipulation and resistance. *Behav. Ecol.* **14**, 80–88 (2003).
- Pellmyr, O. & Thompson, J. N. Multiple occurrences of mutualism in the yucca moth lineage. *Proc. Natl Acad. Sci. USA* **89**, 2927–2929 (1992).
- Sorenson, M. D. & Payne, R. B. A single ancient origin of brood parasitism in African finches: implications for host-parasite coevolution. *Evol. Int. J. Org. Evol.* **55**, 2550–2567 (2001).
- Dale, C., Wang, B., Moran, N. & Ochman, H. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Mol. Biol. Evol.* **20**, 1188–1194 (2003).
- Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman Spektrum, Oxford, 1995).
- Seidler, R. J. & Starr, M. P. Isolation and characterization of host-independent Bdellovibrios. *J. Bacteriol.* **100**, 769–785 (1969).
- Lutzoni, F., Pagel, M. & Reeb, V. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* **411**, 937–940 (2001).
- Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **24**, 403–427 (2000).
- Rosenberg, E. & Varon, M. in *Myxobacteria: Development and Cell Interactions* (ed. Rosenberg, E.) 109–125 (Springer, New York, 1984).
- Wu, S. S. & Kaiser, D. Genetic and functional evidence that Type IV pili are required for social gliding motility in *Myxococcus xanthus*. *Mol. Microbiol.* **18**, 547–558 (1995).
- Shimkets, L. J. Intercellular signaling during fruiting-body development of *Myxococcus xanthus*. *Annu. Rev. Microbiol.* **53**, 525–549 (1999).
- Wireman, J. W. & Dworkin, M. Developmentally induced autolysis during fruiting body formation by *Myxococcus xanthus*. *J. Bacteriol.* **129**, 796–802 (1977).
- Fiegna, F. & Velicer, G. J. Exploitative and hierarchical antagonism in a cooperative bacterium. *PLoS Biol.* **3**, 1980–1987 (2005).
- Velicer, G. J., Kroos, L. & Lenski, R. E. Developmental cheating in the social bacterium *Myxococcus xanthus*. *Nature* **404**, 598–601 (2000).
- Rainey, P. B. & Rainey, K. Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425**, 72–74 (2003).
- Sachs, J. L. & Bull, J. J. Experimental evolution of conflict mediation between genomes. *Proc. Natl Acad. Sci. USA* **102**, 390–395 (2005).
- Velicer, G. J. & Yu, Y. T. Evolution of novel cooperative swarming in the bacterium *Myxococcus xanthus*. *Nature* **425**, 75–78 (2003).
- Velicer, G. J., Kroos, L. & Lenski, R. E. Loss of social behaviors by *Myxococcus xanthus* during evolution in an unstructured habitat. *Proc. Natl Acad. Sci. USA* **95**, 12376–12380 (1998).
- Dugatkin, L. A., Perlin, M., Lucas, J. S. & Atlas, R. Group-beneficial traits, frequency-dependent selection and genotypic diversity: an antibiotic resistance paradigm. *Proc. R. Soc. Lond. B* **272**, 79–83 (2005).
- Greig, D. & Travisano, M. The Prisoner's Dilemma and polymorphism in yeast SUC genes. *Proc. R. Soc. Lond. B* **271** (Suppl. 3), S25–S26 (2004).
- Strassmann, J. E., Zhu, Y. & Queller, D. C. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**, 965–967 (2000).
- Velicer, G. J. Social strife in the microbial world. *Trends Microbiol.* **11**, 330–337 (2003).
- Velicer, G. J., Lenski, R. E. & Kroos, L. Rescue of social motility lost during evolution of *Myxococcus xanthus* in an asocial environment. *J. Bacteriol.* **184**, 2719–2727 (2002).
- Fiegna, F. & Velicer, G. J. Competitive fates of bacterial social parasites: persistence and self-induced extinction of *Myxococcus xanthus* cheaters. *Proc. R. Soc. Lond. B* **270**, 1527–1534 (2003).
- Frank, S. A. Perspective: repression of competition and the evolution of cooperation. *Evol. Int. J. Org. Evol.* **57**, 693–705 (2003).
- Keller, L. *Levels of Selection in Evolution* (Princeton Univ. Press, Princeton, New Jersey, 1999).
- Travisano, M. & Velicer, G. J. Strategies of microbial cheater control. *Trends Microbiol.* **12**, 72–78 (2004).
- Foster, K. R., Shaulsky, G., Strassmann, J. E., Queller, D. C. & Thompson, C. R. L. Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**, 693–696 (2004).
- Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).
- Velicer, G. J., *et al.* Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proc. Natl Acad. Sci. USA* (in the press).
- Vetting, M. W. *et al.* Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **433**, 212–226 (2005).
- Rodríguez, A. M. & Spormann, A. M. Genetic and molecular analysis of *cgIB*, a gene essential for single-cell gliding in *Myxococcus xanthus*. *J. Bacteriol.* **181**, 4381–4390 (1999).
- Orr, H. A. The genetic theory of adaptation: A brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
- Kaiser, D. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proc. Natl Acad. Sci. USA* **76**, 5952–5956 (1979).
- Wall, D., Kolenbrander, P. E. & Kaiser, D. The *Myxococcus xanthus pilQ* (*sgIA*) gene encodes a secretin homolog required for Type IV pilus biogenesis, social motility and development. *J. Bacteriol.* **181**, 24–33 (1999).
- Bretscher, A. P. & Kaiser, D. Nutrition of *Myxococcus xanthus*, a fruiting myxobacterium. *J. Bacteriol.* **133**, 763–768 (1978).

Acknowledgements We thank S. Elena, K. Foster, M. Grbic and our laboratory members for discussions and/or comments on the manuscript. We also thank S. Deiss and H. Keller for technical assistance. This work was partially funded by a grant from the Deutsche Forschungsgemeinschaft.

Author Contributions F.F. (primarily) and G.J.V. (secondarily) performed population-level experiments, Y.-T.N.Y. performed the analysis of mutation function, S.V.K. performed the real-time PCR analysis and all four authors contributed to writing the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.V. (gregory.velicer@tuebingen.mpg.de).

The DNA sequence and biological annotation of human chromosome 1

S. G. Gregory^{1,2}, K. F. Barlow¹, K. E. McLay¹, R. Kaul³, D. Swarbreck¹, A. Dunham¹, C. E. Scott¹, K. L. Howe¹, K. Woodfine⁴, C. C. A. Spencer⁵, M. C. Jones¹, C. Gillson¹, S. Searle¹, Y. Zhou³, F. Kokocinski¹, L. McDonald¹, R. Evans¹, K. Phillips¹, A. Atkinson¹, R. Cooper¹, C. Jones¹, R. E. Hall¹, T. D. Andrews¹, C. Lloyd¹, R. Ainscough¹, J. P. Almeida¹, K. D. Ambrose¹, F. Anderson¹, R. W. Andrew¹, R. I. S. Ashwell¹, K. Aubin¹, A. K. Babbage¹, C. L. Bagguley¹, J. Bailey¹, H. Beasley¹, G. Bethel¹, C. P. Bird¹, S. Bray-Allen¹, J. Y. Brown¹, A. J. Brown¹, D. Buckley³, J. Burton¹, J. Bye¹, C. Carder¹, J. C. Chapman¹, S. Y. Clark¹, G. Clarke¹, C. Clee¹, V. Copley¹, R. E. Collier¹, N. Corby¹, G. J. Coville¹, J. Davies¹, R. Deadman¹, M. Dunn¹, M. Earthrow¹, A. G. Ellington¹, H. Errington¹, A. Frankish¹, J. Frankland¹, L. French¹, P. Garner¹, J. Garnett¹, L. Gay¹, M. R. J. Ghorri¹, R. Gibson¹, L. M. Gilby¹, W. Gillett³, R. J. Glithero¹, D. V. Grafham¹, C. Griffiths¹, S. Griffiths-Jones¹, R. Grocock¹, S. Hammond¹, E. S. I. Harrison¹, E. Hart¹, E. Haugen³, P. D. Heath¹, S. Holmes¹, K. Holt¹, P. J. Howden¹, A. R. Hunt¹, S. E. Hunt¹, G. Hunter¹, J. Isherwood¹, R. James³, C. Johnson¹, D. Johnson¹, A. Joy¹, M. Kay¹, J. K. Kershaw¹, M. Kibukawa³, A. M. Kimberley¹, A. King¹, A. J. Knights¹, H. Lad¹, G. Laird¹, S. Lawlor¹, D. A. Leongamornlert¹, D. M. Lloyd¹, J. Loveland¹, J. Lovell¹, M. J. Lush⁶, R. Lyne¹, S. Martin¹, M. Mashreghi-Mohammadi¹, L. Matthews¹, N. S. W. Matthews¹, S. McLaren¹, S. Milne¹, S. Mistry¹, M. J. F. Moore¹, T. Nickerson¹, C. N. O'Dell¹, K. Oliver¹, A. Palmeiri³, S. A. Palmer¹, A. Parker¹, D. Patel¹, A. V. Pearce¹, A. I. Peck¹, S. Pelan¹, K. Phelps³, B. J. Phillimore¹, R. Plumb¹, J. Rajan¹, C. Raymond³, G. Rouse³, C. Saenphimmachak³, H. K. Sehra¹, E. Sheridan¹, R. Shownkeen¹, S. Sims¹, C. D. Skuce¹, M. Smith¹, C. Steward¹, S. Subramanian³, N. Sycamore¹, A. Tracey¹, A. Tromans¹, Z. Van Helmond¹, M. Wall¹, J. M. Wallis¹, S. White¹, S. L. Whitehead¹, J. E. Wilkinson¹, D. L. Willey¹, H. Williams¹, L. Wilming¹, P. W. Wray¹, Z. Wu³, A. Coulson¹, M. Vaudin¹, J. E. Sulston¹, R. Durbin¹, T. Hubbard¹, R. Wooster¹, I. Dunham¹, N. P. Carter¹, G. McVean⁴, M. T. Ross¹, J. Harrow¹, M. V. Olson³, S. Beck¹, J. Rogers¹ & D. R. Bentley^{1,7}

The reference sequence for each human chromosome provides the framework for understanding genome function, variation and evolution. Here we report the finished sequence and biological annotation of human chromosome 1. Chromosome 1 is gene-dense, with 3,141 genes and 991 pseudogenes, and many coding sequences overlap. Rearrangements and mutations of chromosome 1 are prevalent in cancer and many other diseases. Patterns of sequence variation reveal signals of recent selection in specific genes that may contribute to human fitness, and also in regions where no function is evident. Fine-scale recombination occurs in hotspots of varying intensity along the sequence, and is enriched near genes. These and other studies of human biology and disease encoded within chromosome 1 are made possible with the highly accurate annotated sequence, as part of the completed set of chromosome sequences that comprise the reference human genome.

The sequence of each human chromosome underpins an extremely broad range of biological, genetic and medical studies. Sequence annotation—the process of gathering all of the available information and relating it to the sequence assembly—is essential to develop our understanding of the information stored in human DNA. Initially, there was a strong focus on annotating genes that allowed us to define the genetic information that determines biochemical function and to characterize the functional consequence of genetic aberrations. More recently, we have undertaken systematic identification and annota-

tion of single nucleotide polymorphisms (SNPs) on genomic sequence. This has enabled us to measure the genetic diversity of the genome in geographically distinct population groups, to estimate recombination at a new high-level of resolution, and to identify signals of selection that may reveal new functions encoded in the genome. In parallel, reagents provided by chromosome mapping and sequencing have provided the basis for acquiring additional experimental data: for example, on gene expression and replication timing. These data sets may be used to elucidate the mechanisms that are

¹The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. ²The Duke University Center for Human Genetics, Durham, North Carolina 27708, USA. ³Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, Washington 98195, USA. ⁴King's College London, Department of Medical and Molecular Genetics, Guy's Tower, London SE1 9RT, UK. ⁵Department of Statistics, University of Oxford, Oxford OX1 3TG, UK. ⁶HUGO Gene Nomenclature Committee, The Galton Laboratory, Department of Biology, University College London, London NW1 2HE, UK. ⁷Solexa Ltd, Chesterford Research Park, Little Chesterford, Essex CB10 1XL, UK.

used by the cell to regulate the use of chromosomal sequences—at the level of transcription, epigenetic modification or gross chromosomal behaviour—during replication and cell division.

Chromosome 1 is the largest of the human chromosomes, containing approximately 8% of all human genetic information. Because of its size, we can expect it to be more representative of the human genome than some other chromosomes with respect to genomic landscape and genetic properties. It is medically important: over 350 human diseases are associated with disruptions in the sequence of this chromosome—including cancers, neurological and developmental disorders, and mendelian conditions—for which many of the corresponding genes are unknown. There are also important biological implications of the size of chromosome 1: it is approximately six times longer than the smallest human chromosomes (21, 22 and Y), which raises the question of how all human genetic information is replicated in a coordinated manner before each cell division. This study reports the finished sequence of human chromosome 1, and provides a detailed annotation of the landscape, gene index and sequence variations of the chromosome. Our annotation also brings together information from a wide range of additional genetic and biological studies to describe features such as profiles of recombination, signals of natural selection and replication timing, and their relation to each other along the chromosome sequence. In turn, we show that this level of annotation reveals clues to the location of functionally important sequences that are currently unknown and merit targeted investigation.

Genomic sequence and landscape

We determined the sequence of a set of 2,220 minimally overlapping clones representing the euchromatic portion of chromosome 1 (Supplementary Table S1). The sequence comprises 223,875,858 base pairs (bp) at >99.99% accuracy¹ (Supplementary Table S2); 120,405,438 bp lie in 14 contigs on the short arm (1p) and 103,470,420 bp lie in 13 contigs on the long arm (1q). The sequence reaches telomeric repetitive motifs (TTAGGG)_n on both chromosome arms and pericentromeric alpha-satellite sequence at the proximal end of the short arm (1pcen). There are 18 megabases (Mb) of heterochromatin on 1q adjacent to the centromere that has not been sequenced.

Twenty-six gaps remain after exhaustive screening of bacterial and yeast-derived clone libraries with a combined coverage of 90 genomic equivalents (Supplementary Table S3). Eight gaps are clustered in 1p36 and eight in 1q21.1 (Fig. 1). These regions are GC-rich and contain low-copy repeats, which we believe contribute to the absence of clones in these regions. Seventeen gaps, measured using fluorescent *in situ* hybridization (FISH) of flanking clones to chromosomal DNA, cover a total of 0.8 Mb (data not shown). By aligning the human contigs to the genome sequences of mouse, rat and chimpanzee, we estimated that the remaining nine gaps total 0.53 Mb (Supplementary Table S2). Therefore, the euchromatic fraction of chromosome 1 is 225.2 Mb, and 99.4% is available as finished sequence.

We assessed sequence integrity and completeness by three separate measures. First, all except one of the 2,580 RefSeq genes assigned to chromosome 1 (release number 7; <http://www.ncbi.nlm.nih.gov/RefSeq/>)² are present in the sequence. The missing gene, *RAB7B*, maps to 1q32 in the GB4 radiation hybrid map (<http://www.ncbi.nlm.nih.gov/genemap/>)³ and should lie in gap 23 or 24. Two genes, *IPP* and *PHACTR4*, were only partially represented in the sequence reported here, but have since been completely sequenced (see <http://www.sanger.ac.uk/HGP/sequence/>). Second, we compared the order of 467 chromosome 1 markers in the finished sequence and in the deCODE genetic map⁴ and found no discrepancies. Third, we aligned 32,984 pairs of fosmid clone end sequences to unique positions in the finished sequence and found eleven discordances. Three were sequence misassemblies caused by low-copy repeats, which have been corrected. The remainder are

either deletions in the finished sequence or naturally occurring length polymorphisms. For example, the *GSTM1* gene is absent from 50% of individuals. This gene is present in the reference sequence, but was absent from the fosmid clones mapped to the region.

The G + C, repeat and CpG island content of the chromosome (41%, 48% and 8.9 islands per Mb, respectively) match the genome average⁵ (Supplementary Tables S4 and S5). Areas of high G + C content (Fig. 1c), gene density (Fig. 1d), light Giemsa-staining (Fig. 1a), and SINE (short interspersed element) and LINE (long interspersed element) repeat density (Supplementary Fig. S1a) all correlate⁵. Chromosome 1 has an overall gene density of 14.2 genes per Mb—almost twice the genome average (7.8 genes per Mb) and is, therefore, one of the most gene-dense chromosomes. The 2 Mb light Giemsa-staining region of 1p36.33, adjacent to the 1p telomere, exemplifies a section of extreme sequence content on the chromosome (58.4% G + C, 98 predicted CpG islands, and 104 genes).

Gene index

We curated all available complementary DNA (cDNA) and protein information that provided evidence for gene features, and annotated a total of 3,141 structures (Supplementary Table S6). These are contained in the Vertebrate Genome Annotation (VEGA) database (http://vega.sanger.ac.uk/Homo_sapiens/index.html)⁶. The gene index includes 1,669 known genes, 332 novel coding sequences, 720 novel transcripts, and 420 putative transcripts (defined as described previously⁷), which cover 49.5% of the sequence. We found that 1,189 genes (39%) share overlaps on opposite strands and 655 loci (21%) share overlapping coding regions on the same strand. We also identified 991 pseudogenes, of which 840 are processed, and determined that CpG islands associated with 56% of the known genes and 40% of the novel coding sequences (see Supplementary Methods for details).

We identified evolutionarily conserved regions (ECRs) by

Figure 1 | The genomic landscape of human chromosome 1.

a, Chromosome 1 ideogram according to Francke⁵⁰, showing the differential Giemsa staining pattern. **b**, Sequence scale in intervals of 1 Mb. Note that the correlation between cytogenetic band positions and physical distance is imprecise, owing to varying levels of condensation of different Giemsa bands. **c**, G + C content (on a scale of 30–70%) of 100-kb sequence windows. **d**, Gene density (the number of genes, excluding pseudogenes, per Mb) in 1-Mb sequence windows. **e**, Replication timing ratio (S/G1) at tile-path resolution (horizontal red line denotes the midpoint of replication). **f**, Log of the probability of gene expression at tile-path resolution (horizontal red line denotes the midpoint of log(expression)). **g**, Positions of copy number polymorphisms (CNPs; 1.4 kb–1 Mb in size). **h**, Positions of selected gene families and genes lying within regions of copy number polymorphism (CNP). **i**, Positions of other selected genes of interest on chromosome 1. **j**, Population differences in SNP allele frequency between the CEU, YRI and JPT + CHB HapMap analysis panels²⁴ (see the text for population definitions). The differentiation measure (0–240, on the y axis) is the log-likelihood ratio test statistic. Triangles indicate the most highly differentiated SNPs (see the text for details), and are colour-coded as follows: black, intergenic; purple, intronic; turquoise, untranslated region; red, non-synonymous coding variant. **k**, Haplotype diversity measured for each of the three HapMap panels separately. The blue traces measure the average SNP heterozygosity in windows of 21 SNPs. The horizontal red lines indicate extended haplotypes (defined as the most extreme 1% in terms of length) associated with derived (that is, recent) mutations. The heights of the haplotype bars are arbitrary. **l**, Recombination profile of chromosome 1. The histogram shows recombination rate (cM per Mb) in 100-kb windows. Bars are coloured according to the number of recombination hotspots. Dark red shading indicates the highest level of recombination (5 hotspots per Mb). Dark blue represents regions of least recombination (1 hotspot per Mb). Vertical grey lines in **c–f** and **j–l** represent gaps in the euchromatic sequence of the chromosome. The grey bar located between approximately 121 Mb and 141 Mb shows the position of the centromere and the long-arm heterochromatic block.

alignment of the chromosome 1 sequence to the genome sequences of mouse, rat, zebrafish and two pufferfish species (*Tetraodon nigroviridis* and *Takifugu (Fugu) rubripes*). We found that 10,669 of the 10,971 ECRs conserved in all six genomes overlap with annotated exons, suggesting that exon annotation is at least 97.2% complete (see Supplementary Table S7). The remaining 302 ECRs may represent additional exons without supporting evidence, or highly conserved regulatory or structural elements.

We predicted 459 non-coding RNA (ncRNA) genes or pseudo-genes using the Rfam database of structural RNA alignments (<http://www.sanger.ac.uk/Software/Rfam/>) (Supplementary Table S8). We also identified 22 microRNAs (miRNAs), a class of ncRNAs with a mature length of approximately 22 nucleotides that regulate gene expression by post-transcriptional control, through BLAST analysis of the 1,345 miRNA entries within the miRNA Registry (<http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml>) (Supplementary Table S8).

Sequence duplications

Duplication gives rise to new genetic material that can subsequently diverge and specify novel functions. We analysed the sequence for all repeats ≥ 10 kilobases (kb) in length with $\geq 90\%$ identity, and observed 3.49% intra- and 1.64% inter-chromosomal duplication (Supplementary Fig. S2). A 5-Mb region of 1q21.1 has a complex pattern of intrachromosomal duplication (Fig. 2a). Previously, a bacterial artificial chromosome (BAC) clone derived from 1q21.1

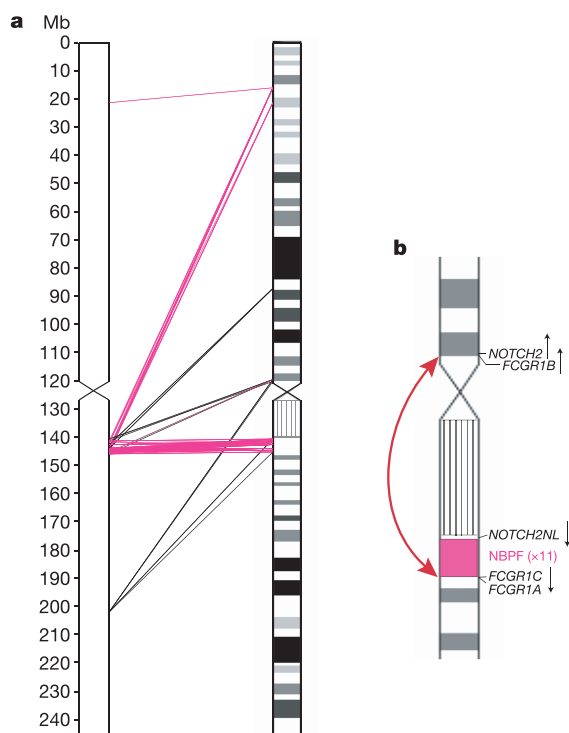


Figure 2 | Segmental duplications involving 1q21.1. **a**, Intrachromosomal segmental duplications within 1q21.1, and those between 1q21.1 and other parts of chromosome 1. The pink lines indicate duplications that incorporate members of the neuroblastoma breakpoint family (NBPF) genes. Three NBPF genes at 1p13 (ref. 9) are not detected in the segmental duplication analysis but are present in the sequence. **b**, Enlarged section of chromosome 1 encompassing the proximal short and long arm (1p13.3–1q23.1). The hatched box is the heterochromatic block on 1q. The positions of the *NOTCH2* and *NOTCH2NL* genes, and the *FCGR1A*, *B* and *C* genes, are shown, and the arrows indicate the direction of their transcription. Eleven members of the NBPF lie in the region of the pink box. The red arrow indicates a suggested pericentric inversion that occurred in the human lineage following duplication of *NOTCH2* and *FCGR1*.

showed FISH signals at 1p36.13 and 1p12, and a broad band of hybridization in 1q21.1 (ref. 8). We found one sequence element at eleven locations in 1q21.1, three locations in 1p36.13, and one location in 1p12. Each copy includes a tandem 1.5-kb repeat array of varying size (≤ 75 kb) that results in exon duplication within different members of the neuroblastoma breakpoint family (NBPF)⁹, so called because one gene (*NBPF1*) was shown to be disrupted by translocation in a neuroblastoma patient. The inter- and intra-genic duplications may foster illegitimate recombination leading to NBPF gene variation⁹.

The NBPF repeat region in human 1q21.1 is flanked by the *NOTCH2NL* gene proximally and the *FCGR1C* and *FCGR1A* genes distally (Fig. 2b). Their homologues—the *NOTCH2* and *FCGR1B* genes—lie together on the other side of the centromere at 1p12, consistent with the occurrence of a pericentric inversion after divergence of the human and chimpanzee lineages¹⁰. *NOTCH2NL* is a truncated copy of *NOTCH2* spanning the 5' end of the gene as far as 8 kb into intron 4. It encodes a functional protein that interacts with neutrophil elastase and has been implicated in hereditary neutropenia¹¹. The *NOTCH2NL* protein contains several of the epidermal growth factor (EGF) repeats found in *NOTCH2*, plus a novel 25-amino-acid carboxy terminus that is required for the interaction with neutrophil elastase¹¹. Our analysis shows that this C-terminal region is derived from the splicing of exon 4 to a region in intron 4.

Paralogous gene pairs result from segmental duplications that diverged by accumulation of mutations at either locus. We found 56 clusters of genes that duplicated after the human–murine divergence (see Supplementary Table S9). The ratio of non-synonymous to synonymous nucleotide substitution rates (K_a/K_s) provides a measure of the rate of divergence in each gene pair. We observed the highest K_a/K_s ratio (1.8) between *SPRR2A* and *SPRR2F* (compared with the chromosome average of 0.4). The *SPRR* genes encode small proline-rich proteins that are primary constituents of the cornified cell envelope—a cross-linked protein scaffold that protects the body from the environment. Many other proteins in this envelope are encoded by genes that cluster with the *SPRR* genes in 1q21.3, and together constitute the epidermal differentiation complex (EDC). The EDC was recently noted as the most rapidly diverging gene cluster in a comparison between human and chimpanzee¹².

Copy number polymorphisms (CNPs) of up to several hundred kilobases occur within phenotypically normal individuals^{13–16}. We positioned CNPs from two CNP databases (<http://paralogy.gs.washington.edu/structuralvariation> and <http://projects.tcag.ca/variation/>) in the chromosome 1 sequence (Fig. 1g and Supplementary Table S10) and localized a subset of gene families within these regions of duplication (Fig. 1h). Genes within CNP regions show structural polymorphism that may lead to disease susceptibility. For example, a *GSTM1* polymorphism may confer an increased cancer risk^{17,18}, and polymorphisms of *FCGR3* have recently been shown to predispose humans to glomerulonephritis¹⁹.

Sequence variation

We mapped 800,653 SNPs from the public databases (dbSNP; release 121) to unique positions in the chromosome 1 sequence. We found 7,917 (1.26%) in protein coding regions, of which 4,471 are non-synonymous and are therefore putative functional variants. We also identified 90 SNPs that introduce premature stop codons in the annotated coding sequence. These mutations would truncate the proteins encoded by 88 genes, 15 of which are associated with genetic diseases and include *COL11A1* in Marshall and Stickler type II syndromes (Online Mendelian Inheritance in Man (OMIM) entry number: 120280), *FY/DARC* in malarial susceptibility (OMIM: 110700), and *UROD* in porphyria cutanea tarda (OMIM: 176100) (see Fig. 1i and Supplementary Table S11). Many of the SNPs on chromosome 1 have been used to determine patterns of genetic variation, providing important new information about molecular

and evolutionary processes that can be annotated along the chromosome sequence (discussed below).

Chromosome recombination

We compared physical and genetic distances between markers from the deCODE genetic map⁴ (Supplementary Fig. S3), and observed high rates of recombination near the telomeres. The lowest rates are near the p-arm centromere (male: 0.04 centimorgans (cM) per Mb; female: 0.77 cM per Mb) and the q-arm heterochromatin (male: 0.04 cM per Mb; female: 0.31 cM per Mb). The sex-averaged recombination rate across the chromosome is 1.13 cM per Mb, equalling the genome average. Recombination is higher in females than males (1.43 versus 0.82 cM per Mb), except at 1ptel and 1qtel (Supplementary Fig. S3).

For a more detailed profile, we used data generated by a recent survey in which the genotypes of 60,000 chromosome 1 SNPs were determined in 269 individuals (as part of the HapMap project²⁰). Recombination rates along the chromosome were estimated using coalescent methodologies²¹, whereby a high level of association between nearby SNP alleles indicates a low level of historic recombination, and *vice versa*. We observed a highly non-random distribution of recombination, with 80% of all recombination occurring in 15% of the sequence, in agreement with previous studies²¹. Peaks of recombination are greater towards the telomeres (Fig. 11). The majority of recombination occurs in hotspots (discrete segments of <2 kb)²¹ of very variable density, but with a trend of higher densities towards the telomeres (vertical red-shaded bars, Fig. 11). In some areas (for example, at 107.5 Mb and 156.5 Mb), a high overall recombination rate is due to a high density of discrete hotspots, whereas elsewhere (for example, at 11.5 Mb and 151 Mb) it is due to a few extremely active hotspots. Elevated recombination is positively correlated with gene density and G + C content. However, further analysis at a finer scale²² reveals that the density of recombination hotspots actually peaks near (within 50 kb), but outside, genes and is suppressed within coding regions. This can be accounted for if double-strand breaks in recombination are accompanied by mutagenesis and, therefore, are sometimes deleterious compared to recombination in non-essential flanking DNA. We also identified an interesting relationship between recombination rate (Fig. 11), expression level (Fig. 1f) and G + C content (Fig. 1c). Contrary to our expectations, we found that higher rates of recombination and G + C content were associated with genes of lower expression (Supplementary Fig. S5).

Natural selection

Natural selection causes altered patterns of genetic variation in populations. Marked differences in the frequency of SNP alleles in one population group relative to another indicate that variants have been selected in one geographically restricted population compared with another. The selected variant is linked with alleles at nearby loci, and evidence for selection may be observed through the existence of extended haplotypes. This effect is influenced by local recombination rate. Therefore, patterns of variation provide a powerful new form of annotation to target searches for sequences that may be important for human fitness. We outline several analyses below.

First, we plotted along chromosome 1 a profile of population-specific differences in SNP allele frequency between populations of Western and Northern European (CEU, Centre d'Etude du Polymorphisme Humain collection (CEPH)/Utah residents from Western and Northern Europe), West African (YRI, Yoruba from Ibadan) or East Asian (JPT + CHB, Japanese from Tokyo + Han Chinese from Beijing) ancestry using information from the HapMap project (Fig. 1j; population samples are described elsewhere²³). The highest peaks identified 68 SNPs (triangles above peaks in Fig. 1j) that provided evidence for geographically restricted selection (see Fig. 1j and Supplementary Table S12). These SNPs included three non-synonymous variants. The best known of these—the *FY**A mutation

in the Duffy gene [OMIM: 110700], which protects against *Plasmodium vivax* malaria—was used as a baseline (log-likelihood ratio test statistic of 150), so that all other 67 SNPs show the same or greater degree of allelic differentiation compared to *FY**A. The two other non-synonymous variants found were in the *ACOT11* gene (a cold-induced thioesterase expressed in adipose tissue and involved in obesity in mice²⁴) and *OR9H1P* (an olfactory receptor that may be a pseudogene). Higher-resolution haplotype analysis of *ACOT11* (Fig. 3) suggests near-fixation for the ancestral haplotype within the YRI population (that is, the ancestral type at all three SNPs—Pro-Asp-Met), whereas the JPT + CHB populations have much greater diversity. This observation may be attributable to strong purifying selection in Africa, relaxed constraints in Asia, and a recent selective sweep in Europe. Most notable among the other SNPs were those clustered in or near *NOTCH2*, *ATP1A1* and *SLC35F3*, and single non-coding SNPs included examples in the cholinergic receptor gene *CHRM3*, a gene encoding a helix-loop-helix transcription factor, and several olfactory receptor genes. We also observed marked allelic

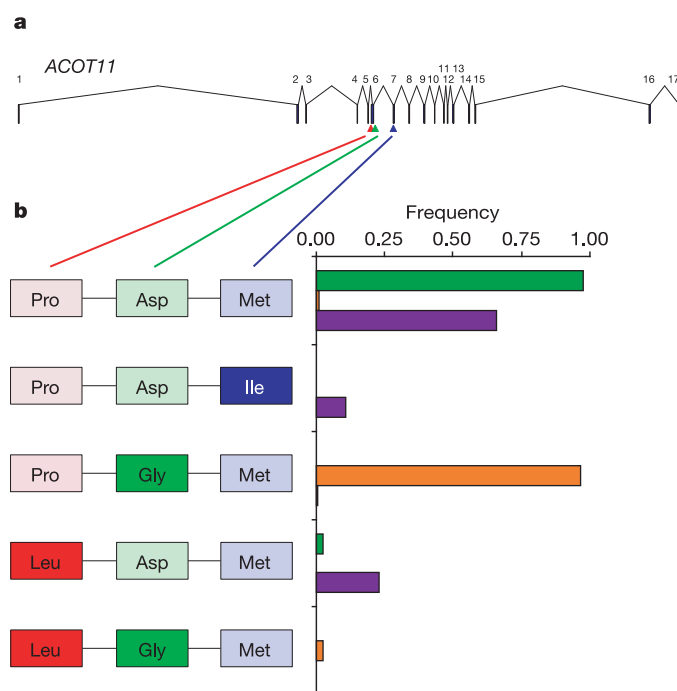


Figure 3 | Population differences in the frequency of non-synonymous haplotypes within the *ACOT11* gene—a cold-induced thioesterase expressed in brown adipose tissue and associated with obesity in mice. **a**, Structure of the *ACOT11* gene, with exons indicated by vertical bars numbered 1–17. Three non-synonymous SNPs, rs2304306, rs1702003 and rs2304305 (red, green and blue arrowheads, respectively), which represent the mutations Pro165Leu, Asp202Gly and Met212Ile, respectively (where the direction of the mutation has been inferred by comparison with the chimpanzee sequences), have been typed across the four HapMap populations. **b**, Different non-synonymous haplotypes in the *ACOT11* gene are shown on the left. Light or dark coloured shading of the boxes containing the amino acids indicate the ancestral or derived mutations at the three sites, respectively. The frequency of each haplotype in each population is shown on the right, colour-coded by population: YRI, green; JPT + CHB, purple; CEU, orange (see the text for population definitions). Note that the population of European origin almost exclusively carries a haplotype with the Asp202Gly mutation—a mutation that is nearly absent from the other populations. In contrast to the situation for most genes, the two Asian populations show the greatest diversity (here grouped together as they have very similar haplotype frequencies), with the African population almost exclusively carrying the ancestral haplotype. These patterns indicate strongly varying selection pressures across the three populations; one possible interpretation being the presence of strong purifying selection in Africa, reduced selection pressures in Asia, and a recent selective sweep in Europe.

differentiation of SNPs associated either with protein coding genes of unknown function, or regions containing no annotated features.

Second, we identified the longest of the extended haplotypes that are associated with the derived (that is, recent) allele for each SNP along the chromosome (horizontal red lines in Fig. 1k)²⁵. These features suggest the occurrence of partial selective sweeps around new beneficial mutations. This analysis revealed several strong candidates for recent adaptive evolution, including some with signals in all populations (for example, at position 92 Mb, which coincides with a high-differentiation SNP in a novel gene (RP11-163M2.4), and at position 35 Mb, which contains the claspin (*CLSPN*) and neurochondrin (*NCDN*) genes). We also found regions with signals in two of the three populations—for example, the extended haplotypes at 50 Mb and 170 Mb in the European (CEU; Fig. 1k) and the Asian (JPT + CHB; Fig. 1k) populations, but not in the African (YRI; Fig. 1k) population. In some cases, these regions are also accompanied by a marked drop in diversity (blue traces in Fig. 1k), which arises because the selected haplotype predominates in the region (see regions 17 Mb, 182 Mb and 221 Mb, where sharp troughs all coincide with high-differentiation SNPs). The correlation of low recombination with the length of the extended haplotype is also evident from co-alignment of these tracks with the recombination rate profile (Fig. 1l). For example, the multiple extended haplotypes at 50 Mb and 170 Mb coincide with the lowest estimated recombination rates on the chromosome.

Replication timing

Human DNA replicates in a distinct temporal manner during the S phase of the cell cycle. It is initiated at replication origins of unknown sequence specificity. Replication timing may be influenced by a chromosome's position within the nucleus^{26,27}, local transcriptional activity²⁸, or base composition^{29–33}, DNA methylation or histone modification^{34,35}. Previously, we surveyed replication timing in lymphoblastoid cells by comparative genomic hybridization of S- or G1-phase nuclear DNA to a microarray of BAC clones at 1-Mb separation across the genome^{33,36}. Chromosome 1 showed the greatest variability in replication timing of any chromosome, indicating that chromosome 1 takes the longest time to replicate. We have performed a higher-resolution study using a microarray of 1,961 overlapping BAC clones (a tile path representing the entire chromosome 1 sequence³⁷). The results confirm the substantial variation in replication timing along the chromosome and suggest some correlations between replication timing and features of the chromosome 1 landscape (see Fig. 1). For example, the distal 45 Mb of the chromosome, which has a high gene-density and G + C content, replicates early in S phase whereas the section of 1p from 55–107 Mb, which contains relatively few genes and a low G + C content, is mostly late in replicating. Linear regression analysis for selected sequence features showed a modest correlation between replication timing and G + C content or SINE content of 0.45 and 0.51, respectively (Supplementary Table S13a). We also tested the relationship between replication timing (Fig. 1e) and gene expression. We obtained expression data points for chromosome 1 genes in 647 of the 1,931 arrayed BAC clones (Fig. 1f) and observed a strong correlation between the probability of gene expression and early replication ($r^2 = 0.83$; Supplementary Fig. S4 and Supplementary Table S13b). This correlation is significantly higher than that obtained from the low-resolution whole-genome study³³, and agrees with the results obtained for equivalent high-resolution analyses on chromosomes 22q and 6 (ref. 36). No significant correlation was observed between replication timing and the level at which a gene is transcribed (Supplementary Table S13c). Replication timing therefore correlates with transcriptional activity, and not necessarily the abundance of specific transcripts in a particular region of the chromosome 1, in agreement with previous reports in *Drosophila*³⁸ and other genome wide studies of the human^{33,36}.

Medical significance

So far, 356 mendelian conditions have been localized to chromosome 1 (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>). These include Parkinson's disease (OMIM: 168600), Alzheimer's disease (OMIM: 104300), several types of Charcot–Marie–Tooth disease (OMIM: 118200, 609260, 118210, 607736, 607791, 607677 and 605253), Gaucher disease (the most common lipid-storage disorder; OMIM: 230800), Stargardt disease 1 (the most common form of inherited juvenile macular degeneration; OMIM: 248200) and the Duffy blood group *FY* gene (malaria susceptibility; OMIM: 110700). The chromosome 1 sequence has contributed to finding the genes involved in 35 mendelian disorders (Supplementary Table S14). Genes for a further 53 diseases remain unknown.

Alterations of chromosome 1—in particular, loss of 1p or gains of 1q—are among the most common chromosomal abnormalities in human cancer. Terminal and interstitial deletions of chromosome 1p occur in as many as 1/5,000 to 1/10,000 live births, and are believed to contribute to mental retardation syndromes. Microarrays consisting of large insert bacterial clones from the chromosome 1 map have been used to characterize genomic rearrangements associated with these diseases and identify candidate genes. This approach has been adopted to elucidate the genes associated with neoplasias such as sarcoma³⁹, meningioma⁴⁰ and pheochromocytoma⁴¹, in addition to developmental phenotypes such as the 1p36 deletion syndrome⁴².

Concluding remarks

At the turn of this century, the finished sequence of the first human chromosomes^{43,44}, and then the draft genome sequence⁵, provided us with a first view of the landscape of the human genome and a partial annotation of genes. Since then, continued efforts to determine the finished sequence of each chromosome has provided a near-complete reference, accurate base-pair scale on which to place all other genetic information. The explosion of parallel investigations using this resource to characterize features of genomic biology, such as sequence variation, disease-causing mutations, recombination, and replication, provide a new level of information with which to annotate chromosome sequences. Our description of chromosome 1 illustrates the importance of continuing the effort to characterize the complexity of information that is stored in finished sequence. Bringing this information together in one place, and providing convenient views of the data and full access to it, is essential to enable us to increase our understanding of genome biology.

METHODS

Mapping, sequencing and sequence analysis. The hierarchical strategy used for the construction of the sequence-ready physical map of chromosome 1, before clone-by-clone sequencing, as well as the tools of the gene annotation pipeline, are as previously described^{6,45}. Manual annotation of gene structures followed the guidelines agreed in the human annotation workshop (HAWK; <http://www.sanger.ac.uk/HGP/havana/hawk.shtml>), and the approved HUGO Gene Nomenclature Committee (<http://www.gene.ucl.ac.uk/nomenclature/>) gene symbols where possible. Known genes are defined as genes with an RNA entry in RefSeq; novel coding sequences are gene structures with experimental evidence and with an open reading frame (ORF); novel transcripts are gene structures with supporting evidence but no obvious ORF; and, putative transcripts are supported gene structures based on alternative species but have no ORF (ref. 7). Protein translations were analysed with InterProScan (<http://www.ebi.ac.uk/InterProScan/>), which was run via the Ensembl protein annotation pipeline to obtain Pfam, Prosite, Prints and Profiles domain matches.

Methods used for shotgun sequencing, finishing strategies, comparative analysis and identification of ECRs, sequence annotation predictions and SNP identification are as previously described⁴⁶, and are also available in Supplementary Methods.

Alignments for inter- and intra-chromosomal duplications were performed with WU-BLASTN (<http://blast.wustl.edu>) using the current sequence assembly of chromosome 1 and National Center for Biotechnology Information (NCBI) build 35 for the rest of the genome. All sequences were repeat-masked with RepeatMasker (<http://repeatmasker.genome.washington.edu>) and low-quality

alignments (E -value $>10^{-30}$; sequence identity $<90\%$; length <80 bp) were discarded. For intrachromosomal duplications, self-matches were discarded. For interchromosomal duplications, the sequence was split into 400-kb segments. Adjacent matches in the same orientation were joined together as described⁴⁷. Only blocks of 10 kb or greater were retained.

SNPs in sequence overlaps were identified using a modification of the SSAHA software⁴⁸. The chromosome 1 SNPs (dbSNP; release 121) were mapped to the sequence assembly of this chromosome (from this study), first with SSAHA and then with Cross-Match.

Replication timing. Replication timing and correlated expression data were generated as previously described³³. Briefly, cells from a human male cultured lymphoblastoid cell line with a normal (46, XY) karyotype (C0202-JAT; European Collection of Cell Cultures (ECCAC) number 94060845) were stained with Hoechst 33258 and flow-sorted into S- and G1-phase fractions. DNA fractions were labelled with dCTP-Cy5 using a random hexamer labelling kit and spin-column-purified before hybridization to a chromosome 1 array.

The overlapping tile-path array of chromosome 1 was constructed by selecting clones from the minimum tiling path of the chromosome⁴² with array hybridization being carried out as described³⁷. The arrays were scanned and images quantified using 'Spot' software⁴⁹. Raw-fluorescence ratios were normalized by dividing each ratio by the mean ratio of all clones and scaled by a value representing the median DNA content of the S-phase fraction calculated from the cell-cycle histogram. The median DNA content of the S-phase fraction was calculated from the mean replication timing ratio reported for chromosome 1 on the 1-Mb-resolution array.

Total RNA was extracted from lymphoblastoid cells and first-strand, second-strand cDNA synthesis was performed on 10 μ g of total RNA using 100 pmol of a high-performance liquid chromatography (HPLC)-purified T7-(T)₂₄ primer. Amplified, biotinylated complementary RNA was then produced with an *in vitro* transcription labelling reaction. Samples with a yield greater than 40 μ g of cRNA were subsequently hybridized to Affymetrix U133A oligonucleotide arrays. Hybridization was performed at 45 °C for 16 h. Arrays were washed and stained with streptavidin-phycoerythrin. Signal amplification was performed using a biotinylated anti-streptavidin antibody following the recommended Affymetrix protocol for high-density chips. Scans were carried out on a GeneArray scanner. The fluorescence intensities of scanned arrays were analysed with Affymetrix GeneChip software. The Affymetrix Microarray Suite 5.0 was used for the quantification of gene expression levels. Global scaling was applied to the data to adjust the average recorded intensity to a target intensity of 100. Quantification data was exported from Affymetrix Microarray Suite 5.0 into Microsoft Office Excel for further analysis. Presence or absence of gene expression was determined by a 'present' call in any of the oligonucleotides representing a gene, as determined by Affymetrix Microarray Suite 5.0.

Methods for the determination of genome parameters associated with replication timing analysis are available in Supplementary Methods, and the analysis of segmentation data was carried out as previously described³³.

Received 24 December 2005; accepted 13 March 2006.

- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
- Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
- Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Ashurst, J. L. *et al.* The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res.* **33**, D459–D465 (2005).
- Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
- Weise, A., Starke, H., Mrasek, K., Claussen, U. & Liehr, T. New insights into the evolution of chromosome 1. *Cytogenet. Genome Res.* **108**, 217–222 (2005).
- Vandepoele, K., Van Roy, N., Staes, K., Speleman, F. & Van Roy, F. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol. Biol. Evol.* **22**, 2265–2274 (2005).
- Maresco, D. L. *et al.* Localization of FCGR1 encoding Fc γ receptor class I in primates: molecular evidence for two pericentric inversions during the evolution of human chromosome 1. *Cytogenet. Cell Genet.* **82**, 71–74 (1998).
- Duan, Z. *et al.* A novel Notch protein, N2N, targeted by neutrophil elastase and implicated in hereditary neutropenia. *Mol. Cell. Biol.* **24**, 58–70 (2004).
- Chimp Sequencing Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
- Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- lafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
- Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
- Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
- Strange, R. C. *et al.* The human glutathione S-transferases: a case-control study of the incidence of the GST1 O phenotype in patients with adenocarcinoma. *Carcinogenesis* **12**, 25–28 (1991).
- van Poppel, G., de Vogel, N., van Balderen, P. J. & Kok, F. J. Increased cytogenetic damage in smokers deficient in glutathione S-transferase isozyme μ . *Carcinogenesis* **13**, 303–305 (1992).
- Aitman, T. J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
- The International HapMap Consortium, The International HapMap Project. *Nature* **426**, 789–796 (2003).
- McVean, G. A. T. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
- Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. Genetics: A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
- The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- Adams, S. H. *et al.* BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: cloning, organization of the human gene and assessment of a potential link to obesity. *Biochem. J.* **360**, 135–142 (2001).
- Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
- Ferreira, J., Paoletti, G., Ramos, C. & Lamond, A. I. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J. Cell Biol.* **139**, 1597–1610 (1997).
- Schermelleh, L., Solovei, I., Zink, D. & Cremer, T. Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res.* **9**, 77–80 (2001).
- Holmquist, G. P. Role of replication time in the control of tissue-specific gene expression. *Am. J. Hum. Genet.* **40**, 151–173 (1987).
- Hassan, A. B. & Cook, P. R. Does transcription by RNA polymerase play a direct role in the initiation of replication? *J. Cell Sci.* **107**, 1381–1387 (1994).
- Hassan, A. B., Jackson, D. A., Cook, P. R., Errington, R. J. & White, N. S. Replication and transcription sites are colocalized in human cells. *J. Cell Sci.* **107**, 425–434 (1994).
- Gilbert, D. M. Making sense of eukaryotic DNA replication origins. *Science* **294**, 96–100 (2001).
- Gilbert, D. M. Replication timing and metazoan evolution. *Nature Genet.* **32**, 336–337 (2002).
- Woodfine, K. *et al.* Replication timing of the human genome. *Hum. Mol. Genet.* **13**, 191–202 (2004).
- Cimbora, D. M. *et al.* Long-distance control of origin choice and replication timing in the human β -globin locus are independent of the locus control region. *Mol. Cell. Biol.* **20**, 5581–5591 (2000).
- Schübeler, D. *et al.* The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**, 1263–1271 (2004).
- Woodfine, K. *et al.* Replication timing of human chromosome 6. *Cell Cycle* **4**, 172–176 (2005).
- Fiegler, H. *et al.* DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosom. Cancer* **36**, 361–374 (2003).
- Schübeler, D. *et al.* Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nature Genet.* **32**, 438–442 (2002).
- Kresse, S. H. *et al.* Mapping and characterization of the amplicon near APOA2 in 1q23 in human sarcomas by FISH and array CGH. *Mol. Cancer* **4**, 39 (2005).
- Buckley, P. G. *et al.* Comprehensive DNA copy number profiling of meningioma using a chromosome 1 tiling path microarray identifies novel candidate tumor suppressor loci. *Cancer Res.* **65**, 2653–2661 (2005).
- Jarbo, C. *et al.* Detailed assessment of chromosome 22 aberrations in sporadic pheochromocytoma using array-CGH. *Int. J. Cancer* **118**, 1159–1164 (2005).
- Redon, R. *et al.* Tiling path resolution mapping of constitutional 1p36 deletions by array-CGH: contiguous gene deletion or "deletion with positional effect" syndrome? *J. Med. Genet.* **42**, 166–171 (2005).
- Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
- Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
- Bentley, D. R. *et al.* The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* **409**, 942–943 (2001).

46. Mungall, A. J. *et al.* The DNA sequence and analysis of human chromosome 6. *Nature* **425**, 805–811 (2003).
47. Cheung, J. *et al.* Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**, R25 (2003).
48. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
49. Jain, A. N. *et al.* Fully automatic quantification of microarray image data. *Genome Res.* **12**, 325–332 (2002).
50. Francke, U. Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet. Cell Genet.* **65**, 206–218 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank the numerous collaborators who have contributed experimental data to the construction of the physical map and

assembly of the finished sequence, the EMBL and Ensembl database teams at the European Bioinformatics Institute, S. Povey, E. A. Bruford, T. A. Eyre, V. K. Khodiyar, R. C. Lovering, K. M. B. Sneddon, T. P. Sneddon, C. C. Talbot Jr and M. W. Wright at the HUGO Gene Nomenclature Committee for assignment of official gene symbols, and T. Furey for data mining from the UCSC database. Work at the Sanger Institute was funded by the Wellcome Trust, work at the University of Washington was funded by the NIH, and work at HUGO was funded by the NIH and the MRC.

Author Information The updated human chromosome 1 sequence can be accessed through GenBank accession number NC_000001. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.G. (sgregory@chg.duhs.duke.edu).

LETTERS

Suppression of dwarf galaxy formation by cosmic reionization

J. Stuart B. Wyithe¹ & Abraham Loeb²

A large number of faint galaxies, born less than a billion years after the Big Bang, have recently been discovered^{1–6}. Fluctuations in the distribution of these galaxies contributed to a scatter in the ionization fraction of cosmic hydrogen on scales of tens of megaparsecs, as observed along the lines of sight to the earliest known quasars^{7–9}. Theoretical simulations predict that the formation of dwarf galaxies should have been suppressed after cosmic hydrogen was reionized^{10–13}, leading to a drop in the cosmic star-formation rate¹⁴. Here we report evidence for this suppression. We show that the post-reionization galaxies that produced most of the ionizing radiation at a redshift $z \approx 5.5$ must have had a mass in excess of $\sim 10^{10.9 \pm 0.5}$ solar masses ($M_\odot$) or else the aforementioned scatter would have been smaller than observed. This limiting mass is two orders of magnitude larger than the galaxy mass that is thought to have dominated the reionization of cosmic hydrogen ($\sim 10^8 M_\odot$). We predict that future surveys with space-based infrared telescopes will detect a population of smaller galaxies that reionized the Universe at an earlier time, before the epoch of dwarf galaxy suppression.

The lines of sight towards the earliest known quasars at redshifts $z > 6$ show a rapid transition in the ionization state of the intergalactic medium (IGM), potentially marking the end of the reionization epoch^{15–18}. It is thought that the dominant source population to have triggered reionization included dwarf galaxies with virial temperatures above the hydrogen cooling threshold ($T_{\text{vir}} \approx 10^4$ K). The re-ionization of cosmic hydrogen resulted in heating of the IGM to $\sim 10^4$ K, which drastically increased the minimum virial temperature of new galaxies^{12,13} to a value as high as $T_{\text{vir}} \approx 10^5$ K. Reionization is therefore expected to be followed by a depletion of low-mass galaxies and hence a decrease in the global star-formation rate¹⁴. The detection of this anticipated suppression at $z \approx 6$ would provide important evidence for a late reionization epoch.

In addition to luminous quasars, starburst galaxies with star-formation rates in excess of $\sim 0.1 M_\odot \text{ yr}^{-1}$ and dark-matter haloes⁶ of $\sim 10^9$ – $10^{10} M_\odot$ have recently been discovered^{1–6,19–22} at $z \approx 5$ – 6 . These sources contributed to the ionizing background radiation at the end of the reionization epoch. Luminous Lyman α emitters are routinely identified through continuum dropout and narrow-band imaging techniques^{3,4,20}. However, to study fainter sources which were potentially responsible for reionization, spectroscopic searches have been undertaken near the critical curves of lensing galaxy clusters^{21–23}, where gravitational magnification enhances the flux sensitivity. A comparison between the findings of these deep surveys and the statistics of bright Ly α emitters²² provides a preliminary hint for the depletion of low-mass galaxies, as expected from galaxy formation in a photo-ionized IGM^{12,13}. However, this inference relies on a comparison between heterogeneous data sets based on different methods of discovery.

Figure 1a shows data on the luminosity function of galaxies at

$z \approx 5.5$ – 6 from the Hubble Space Telescope Ultra-Deep Field⁴. We first use this data to constrain the physical characteristics of high redshift galaxies (see Supplementary Information 1). We model starburst galaxies based on a simple prescription for their continuum luminosity²⁴ combined with the Sheth–Tormen²⁵ mass function of galaxy haloes. The two free parameters in this model are the starburst lifetime (t_{lt}) and the mass fraction of virialized baryons which are converted into stars in each galaxy (f_{star}).

Figure 1b shows contours of the joint *a posteriori* probability distribution $d^2P/d(\log f_{\text{star}})d(\log t_{\text{lt}})$ (see Supplementary Information 2 for a description). The best-fit luminosity function is plotted over the data in Fig. 1a (blue) with the favoured values of t_{lt} and f_{star} listed. The other contours in Fig. 1b (grey) show the virial temperature corresponding to the lowest luminosity bin, labelled by $\log_{10}(T_{\text{vir}})$. The best-fit luminosity function implies that the faintest observed galaxies have a virial temperature $\sim 10^5$ K, and that starburst galaxies at $z \approx 6$ have duty cycles of $t_{\text{lt}}/t_{\text{H}} \approx 10\%$ (where t_{H} is the Hubble time), combined with star-formation efficiencies of $f_{\text{star}} \approx 10$ – 15% . Interestingly, a duty cycle of $t_{\text{lt}}/t_{\text{H}} \approx 0.1$ corresponds to the dynamical time at the virial overdensity of 200, that is, the time it takes the virialized gas to settle to the centre of a collapsing galaxy. For the purpose of understanding the process of galaxy formation at high redshifts and the way it shapes reionization, it is crucial to know whether star formation was suppressed in dwarf galaxies near the hydrogen-cooling threshold. With a virial temperature of $T_{\text{vir}} \approx 10^4$ K, these galaxies had a total (halo) mass of $\sim 10^8 M_\odot$ at $z \approx 6$. Unfortunately, present detection limits are insufficient to probe the existence of such galaxies.

We overcome the limitation of direct detection of faint galaxies by making use of the fact that the absorption spectra of high-redshift quasars reveal large fluctuations in the IGM ionization state on very large scales^{7–9,17} (tens to hundreds of megaparsecs). While observations of the galaxy luminosity function are flux-limited, the fluctuations in the ionizing background are sensitive to contributions from galaxies of all luminosities. A measurement of the fluctuations in the ionizing background therefore provides a powerful probe of the properties of the galaxies generating the ionizing background radiation at the end of the reionization era, even if those galaxies themselves lie below current detection limits. The background fluctuations can be inferred from observed fluctuations in the effective Ly α optical depth, τ_{eff} .

Fan *et al.*⁹ have measured average values for the effective optical depths of $\tau_{\text{eff}} = 2.5, 2.6, 3.2$ and 4.0 , within redshift bins of width $\Delta z = 0.15$ at $z = 5.25, 5.45, 5.65$ and 5.85 , respectively. At these redshifts the scatter in τ_{eff} among the different lines of sight is $\sigma_\tau = 0.5, 0.6, 0.8$ and 0.8 , respectively. The dominant uncertainty in individual estimates of τ_{eff} is the level of unabsorbed quasar continuum, which introduces noise⁹ in τ_{eff} of ± 0.05 . This observational uncertainty must be subtracted (in quadrature) from σ_τ to obtain an

¹University of Melbourne, Parkville, Victoria 3010, Australia. ²Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

estimate of the intrinsic scatter in τ_{eff} . Following this operation we find fractional fluctuations $\delta_\tau \equiv \sigma_\tau / \tau_{\text{eff}}$ of (0.20 ± 0.03) , (0.23 ± 0.03) , (0.25 ± 0.03) and (0.20 ± 0.03) at the above sequence of redshifts. The corresponding co-moving mean free paths of ionizing photons are $R_{\text{mfp}} = (39^{+16}_{-10})\text{Mpc}$, $(41^{+21}_{-15})\text{Mpc}$, $(27^{+21}_{-10})\text{Mpc}$ and $(19^{+10}_{-15})\text{Mpc}$ respectively⁹. Here we have corrected the frequency-averaged ionization cross-section for a spectrum corresponding to a star-forming galaxy²⁶. At the highest redshift under consideration, some lines of sight contain blank absorption troughs, making their estimated value of τ_{eff} a lower limit. At $z \approx 5.85$ there is no measured lower limit for R_{mfp} , and so we conservatively set it to 3 Mpc (below the typical value at $z \approx 6$).

We have calculated the predicted scatter in τ_{eff} from a combination of cosmic variance among different regions and Poisson noise in the number of ionizing sources, based on an extension of a detailed model published elsewhere²⁷. A description of this model is provided in Supplementary Information 3, and only a summary of its main features is given here. Fluctuations in both the ionizing background

intensity J and the density field around their mean values, contribute to fluctuations in τ_{eff} . Within our model, the ionizing background is smoothed on the scale R_{mfp} , because any given point in the IGM is exposed to all sources within a volume $V_{\text{mfp}} = (4\pi/3)R_{\text{mfp}}^3$ around it.

The model combines four major sources of fluctuations: (1) Poisson noise in the number of sources contained within a volume V_{mfp} . (2) Fluctuations in the value of J within volumes of V_{mfp} resulting from delayed (or enhanced) structure formation in underdense (overdense) regions²⁸. In evaluating this quantity we include two contributions to the derivative dJ/dz arising from the evolution of the star-formation rate (which at first we take to be proportional to the derivative of the mass fraction of baryons that were assembled into galaxies) and the evolution of R_{mfp} . (3) Fluctuations in Ly α transmission due to fluctuations in the density contrast smoothed on the scale R_{mfp} . (4) Fluctuations in the transmission due to small-scale density fluctuations along the line of sight through a region with constant J . These latter fluctuations are computed on the basis of the results of numerical simulations²⁹.

We have compared the predictions of the above model with observational data as a function of the lifetime t_{lt} and the minimum virial temperature T_{min} of the lowest-mass galaxies that make a major contribution to the ionizing background. Contours of the resulting distribution $d^2P/d(\log t_{\text{lt}}d(\log T_{\text{min}}))$ (see Supplementary Information 4 for details) evaluated at $z \approx 5.45$ (blue contours) and $z \approx 5.65$ (red contours) are plotted in Fig. 1c. The corresponding marginalized differential probability distributions $dP/d(\log T_{\text{min}})$ are plotted in Fig. 1d (solid blue and red lines). In addition we plot $dP/d(\log T_{\text{min}})$ at $z \approx 5.25$ (dashed blue line) and $z \approx 5.85$ (dashed red line) in Fig. 1d. Intriguingly, the results at $z \lesssim 5.65$ favour $T_{\text{min}} \approx 10^5\text{K}$, while results at $z \approx 5.85$ favour $T_{\text{min}} \approx 10^4\text{--}10^5\text{K}$, though we note that $\sigma_\tau(z = 5.85)$ is a lower limit. This trend is expected if reionization completed at $z \approx 6\text{--}7$. Indeed, the mini-haloes ($T_{\text{vir}} < 10^4\text{K}$) that provide the collapsed gas reservoir for star formation in galaxies above the hydrogen-cooling threshold are thought to be photo-evaporated³⁰ on a timescale of $\sim 0.2t_{\text{H}}$, which corresponds to $\Delta z \approx 1$. Thus, if reionization completed at $z \approx 6.5\text{--}7$, we would only expect to observe the strong effect of dwarf galaxy suppression at $z \lesssim 5.5\text{--}6$.

We also plot the combined differential and cumulative distributions using data from the three redshift bins below $z \approx 5.75$ (black lines), and find values of $T_{\text{min}} \approx 10^{5.5 \pm 0.3}\text{K}$ (68%), with $T_{\text{min}} \leq 10^4\text{K}$ ruled out at a (statistical) confidence level of $>99\%$. The above constraints have conservatively assumed a prior probability distribution for T_{min} that extends down to $T_{\text{min}} = 10^3\text{K}$. On the other hand, the ratio between the probabilities of having $10^4\text{K} < T_{\text{min}} < 10^5\text{K}$ and $10^5\text{K} < T_{\text{min}} < 10^6\text{K}$ is independent of any assumed lower cut-off. We find this ratio to be ~ 0.05 . A limiting virial temperature of $T_{\text{min}} \geq 10^5\text{K}$ is an order of magnitude above the threshold for star formation triggered by atomic hydrogen cooling. However, $T_{\text{min}} \geq 10^5\text{K}$ is consistent with galaxy formation in an IGM that was heated by ultraviolet photo-ionization at a higher redshift, thus preventing starbursts in low-mass galaxies from dominating the global star-formation rate by $z \approx 5.5$.

Our analysis so far has assumed for simplicity that the star-formation rate is proportional to the derivative of the collapsed fraction in haloes with $T_{\text{vir}} > T_{\text{min}}$. However, starburst episodes are believed to be triggered by galaxy mergers. We have therefore repeated the above calculation, replacing the derivative of the collapsed fraction with a star-formation model¹⁴ that includes the contribution from mergers of haloes with an initial $T_{\text{vir}} < T_{\text{min}}$ (leading to $T_{\text{vir}} > T_{\text{min}}$ after the merger) and the contribution from mergers of galaxies with an initial $T_{\text{vir}} > T_{\text{min}}$ (see Supplementary Information 5 for details). The results are shown in Supplementary Fig. 1. This alternative estimate of the star-formation rate also leads to a preferred value of $T_{\text{min}} \approx 10^{5.5 \pm 0.3}\text{K}$, with $T_{\text{min}} \geq 10^4\text{K}$ at the $\geq 99\%$ confidence level.

In the local Universe, star formation within galaxies with stellar

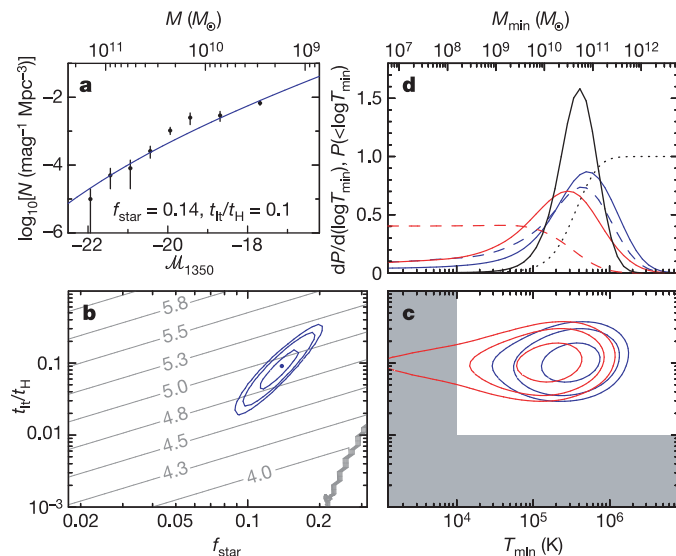


Figure 1 | Constraints on the star-formation efficiency (f_{star}), duty cycle ($t_{\text{lt}}/t_{\text{H}}$) and minimum virial temperature (T_{min}) of galaxies at $z \approx 5.5\text{--}6$. **a**, The luminosity function inferred from the Hubble Ultra-Deep Field⁴, together with the best-fit solution for a luminosity function model based on the Sheth–Tormen²⁵ mass function (with luminosity proportional to halo mass). M_{1350} is the galaxy's absolute AB magnitude at a rest-frame wavelength of 1,350 Å. The upper axis shows the halo mass corresponding to M_{1350} for the best fit. Error bars are 1σ . **b**, Contours (at 64, 26 and 14% of the maximum likelihood) of the joint *a posteriori* probability distribution: $d^2P/d(\log f_{\text{star}}d(\log t_{\text{lt}}))$. For comparison, the grey contours show the virial temperature of haloes—labelled using $\log_{10}(T_{\text{vir}})$ —corresponding to the lowest-luminosity bin in the Hubble Ultra-Deep Field. **c**, The joint probability distributions for $\log(t_{\text{lt}}/t_{\text{H}})$ and $\log(T_{\text{min}})$, which include the constraint on $t_{\text{lt}}/t_{\text{H}}$ from the galaxy luminosity function. The contours represent values at 64, 26 and 14% of the maximum likelihood, and are shown for observations at $z \approx 5.45$ (blue contours) and $z \approx 5.65$ (red contours). This distribution includes constraints on t_{lt} from the galaxy luminosity function. However, assuming the most likely value of R_{mfp} , we find that $T_{\text{min}} \geq 10^5\text{K}$ at $z \lesssim 5.5$ in all cases where the duty cycle of the ultraviolet-bright phase of galaxies is greater than 1% (corresponding to the lifetime of massive stars). **d**, Differential probability distributions for $\log(T_{\text{min}})$ at $z \approx 5.25$ (dashed blue lines), $z \approx 5.45$ (blue lines), $z \approx 5.65$ (red lines) and $z \approx 5.85$ (dashed red lines). We also show combined constraints using the three lowest-redshift bins (black lines). For comparison, the upper axis shows the corresponding masses at $z = 5.5$. In this case the solid and dotted lines correspond to differential and cumulative distributions. Throughout this Letter, we adopt the latest values for the cosmological parameters as inferred from the Wilkinson Microwave Anisotropy Probe data³³.

masses below $\sim 10^{10} M_{\odot}$ is suppressed as a result of supernova-driven winds which expel gas from their shallow potential wells^{31,32}. At high redshifts the lifetime of massive stars could be comparable to the halo dynamical time and this feedback may not operate as effectively. Nevertheless, it is important to explore the effect of such a feedback on fluctuations in the ionizing background and hence on our conclusions regarding the suppression of dwarf galaxy formation in the reionized IGM. We have repeated our analysis of the galaxy luminosity function, with an allowance for a suppression of star formation in galaxies below a critical mass (which is treated as a free parameter). Details of the required modifications and the results are described in Supplementary Information 6 and Supplementary Fig. 2. Allowing for feedback in high-redshift galaxies leads to estimates of $f_{\text{star}} \approx 0.5$, with a duty cycle that is larger than $t_{\text{lt}}/t_{\text{H}} \approx 0.5$. The best-fit value for the critical mass scale below which supernovae feedback operates is $\sim 2 \times 10^{10} M_{\odot}$, or $T_{\text{vir}} \approx 10^5$ K.

Estimates of the scatter in the ionizing background due to Poisson noise in the number of galaxies and due to fluctuations in the star-formation rate as a result of delayed or enhanced structure formation must also be modified to account for possible supernova feedback. We find that if supernova feedback operates at high redshift in the same way as observed at low redshift, then the most likely value is $T_{\text{min}} \approx 10^{5.5+0.3}_{-0.4}$ K in agreement with our previous calculation, while $T_{\text{min}} \leq 10^4$ K is ruled out at the $\geq 95\%$ level. Note that the critical mass scales derived for supernovae feedback and M_{min} are similar, although they were derived from different data sets and astrophysical considerations. However, our results imply that a sharp cut-off due to reionization is required to explain the observed scatter in the effective optical depth of the IGM, in addition to any suppression resulting from supernovae feedback.

The observed star-formation rate in the post-reionization era is on the borderline of being sufficient to reionize the Universe⁴. The value of $T_{\text{min}} \approx 10^5$ K at $z \approx 5.5$ happens to describe the faintest galaxies in the Hubble Ultra-Deep Field. We therefore predict that star formation is suppressed in galaxies just below the current limit of galaxy surveys at $z \approx 6$. Future searches at higher redshifts with the James Webb Space Telescope or large-aperture ground-based infrared telescopes should find a population of galaxies with $T_{\text{vir}} \approx 10^4$ K and detect an enhanced star-formation rate prior to completion of reionization¹⁴.

Received 19 January; accepted 20 March 2006.

- Rhoads, J. E. & Malhotra, S. Ly α emitters at $z = 5.7$. *Astrophys. J. Lett.* **563**, 5–8 (2001).
- Shimasaku, K. *et al.* Subaru Deep Survey. IV. Discovery of a large-scale structure at redshift ~ 5 . *Astrophys. J. Lett.* **586**, 111–114 (2003).
- Bouwens, R. J. *et al.* Galaxies at $z = 7$ –8: z_{850} -dropouts in the Hubble Ultra Deep Field. *Astrophys. J. Lett.* **616**, 79–82 (2004).
- Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P. & Franx, M. Galaxies at $z \sim 6$: The rest-frame UV luminosity function and luminosity density from 506 UDF, UDF-PS, and goods I-dropouts. Preprint at (<http://arXiv.org/astro-ph/0509641>) (2005).
- Ouchi, M. *et al.* Subaru Deep Survey. VI. A census of Lyman break galaxies at $z = 4$ and 5 in the Subaru deep fields: Clustering properties. *Astrophys. J.* **611**, 685–702 (2004).
- Stark, D. P. & Ellis, M. Searching for the sources responsible for cosmic reionization: Probing the redshift range $7 < z < 10$ and beyond. Preprint at (<http://arXiv.org/astro-ph/0508123>) (2005).
- Djorgovski, S. G. in *Proc. Tenth Marcel Grossmann Meeting on General Relativity* (eds Novello, M., Perez-Bergliaffa, S. & Ruffini, R.) (World Scientific, Singapore, in the press); preprint at (<http://arXiv.org/astro-ph/0409378>) (2004).
- Djorgovski, S. G., Bogosavljevic, M. & Mahabal, A. Quasars as probes of late reionization and early structure formation. *New Astron. Rev.* **50**, 140–145 (2006).
- Fan, X. *et al.* Constraining the evolution of the ionizing background and the epoch of reionization with $z \sim 6$ quasars. II: Analysis using a sample of 19 quasars. Preprint at (<http://arXiv.org/astro-ph/0512082>) (2005).
- Efstathiou, G. Suppressing the formation of dwarf galaxies via photoionization. *Mon. Not. R. Astron. Soc.* **256**, 43–47 (1992).
- Quinn, T., Katz, N. & Efstathiou, G. Photoionization and the formation of dwarf galaxies. *Mon. Not. R. Astron. Soc.* **278**, L49–L54 (1996).
- Thoul, A. A. & Weinberg, D. H. Hydrodynamic simulations of galaxy formation. II. Photoionization and the formation of low-mass galaxies. *Astrophys. J.* **465**, 608–616 (1996).
- Dijkstra, M., Haiman, Z., Rees, M. J. & Weinberg, D. H. Photoionization feedback in low-mass galaxies at high redshift. *Astrophys. J.* **601**, 666–675 (2004).
- Barkana, R. & Loeb, A. Identifying the reionization redshift from the cosmic star formation rate. *Astrophys. J.* **539**, 20–25 (2000).
- Fan, X. *et al.* A survey of $z > 5.7$ quasars in the Sloan Digital Sky Survey III: Discovery of five additional quasars. *Astron. J.* (in the press); preprint at (<http://arXiv.org/astro-ph/0405138>) (2005).
- Pentericci, L. *et al.* VLT optical and near-infrared observations of the $z = 6.28$ quasar SDSS J1030+0524. *Astron. J.* **123**, 2151–2158 (2002).
- White, R. L., Becker, R. H., Fan, X. & Strauss, M. A. Probing the ionization state of the Universe at $z > 6$. *Astron. J.* **126**, 1–14 (2003).
- Wyithe, J. S. B. & Loeb, A. A large neutral fraction of cosmic hydrogen a billion years after the Big Bang. *Nature* **427**, 815–817 (2004).
- Rhoads, J. E. *et al.* Spectroscopic confirmation of three redshift $z = 5.7$ Ly α emitters from the large-area Lyman alpha survey. *Astron. J.* **125**, 1006–1013 (2003).
- Bunker, A., Stanway, E., Ellis, R. & McMahon, R. G. The star formation rate of the Universe at $z \sim 6$ from the Hubble Ultra-Deep Field. *Mon. Not. R. Astron. Soc.* **355**, 374–384 (2004).
- Ellis, R., Santos, M., Kneib, J.-P. & Kuijken, K. A faint star-forming system viewed through the lensing cluster Abell 2218: First light at $z \sim 5.6$? *Astrophys. J. Lett.* **560**, 119–122 (2001).
- Santos, M., Ellis, R., Kneib, J.-P., Richard, J. & Kuijken, K. The abundance of low-luminosity Ly α emitter at high redshift. *Astrophys. J.* **606**, 683–701 (2004).
- Kneib, J.-P., Ellis, R. S., Santos, M. R. & Richard, J. A probable $z \sim 7$ galaxy strongly lensed by the rich cluster A2218: Exploring the Dark Ages. *Astrophys. J.* **607**, 697–703 (2004).
- Loeb, A., Barkana, R. & Hernquist, L. Was the Universe reionized at redshift 10? *Astrophys. J.* **620**, 553–558 (2005).
- Sheth, R. K., Mo, H. J. & Tormen, G. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Mon. Not. R. Astron. Soc.* **323**, 1–12 (2001).
- Leitherer, C. *et al.* Starburst99: Synthesis models for galaxies with active star formation. *Astrophys. J. Suppl.* **123**, 3–40 (1999).
- Wyithe, J. S. B. & Loeb, A. Cosmic variance in the transparency of the intergalactic medium after reionization. Preprint at (<http://arXiv.org/astro-ph/0508604>) (2006).
- Barkana, R. & Loeb, A. Unusually large fluctuations in the statistics of galaxy formation at high redshift. *Astrophys. J.* **609**, 474–481 (2004).
- Lidz, A., Oh, S. P. & Furlanetto, S. R. Have we detected patchy reionization in quasar spectra? *Astrophys. J. Lett.* **639**, 47–50 (2006).
- Barkana, R. & Loeb, A. Effective screening due to minihalos during the epoch of reionization. *Astrophys. J.* **578**, 1–11 (2002).
- Dekel, A. & Woo, J. Feedback and the fundamental line of low-luminosity low-surface-brightness/dwarf galaxies. *Mon. Not. R. Astron. Soc.* **344**, 1131–1144 (2003).
- Kauffmann, G. *et al.* The dependence of star formation history and internal structure on stellar mass for 10^5 low-redshift galaxies. *Mon. Not. R. Astron. Soc.* **341**, 54–69 (2003).
- Spergel, D. N. *et al.* Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for cosmology. *Astrophys. J.* (submitted); preprint at (http://lambda.gsfc.nasa.gov/product/map/dr2/pub_papers/threeyear/parameters/wmap_3yr_param.pdf) (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by grants from ARC, NSF and NASA. We acknowledge helpful discussions with A. Lidz regarding this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.L. (aloeb@cfa.harvard.edu) or S.W. (swyithe@physics.unimelb.edu.au).

An aluminium nitride light-emitting diode with a wavelength of 210 nanometres

Yoshitaka Taniyasu¹, Makoto Kasu¹ & Toshiaki Makimoto¹

Compact high-efficiency ultraviolet solid-state light sources¹—such as light-emitting diodes (LEDs) and laser diodes—are of considerable technological interest as alternatives to large, toxic, low-efficiency gas lasers and mercury lamps. Microelectronic fabrication technologies and the environmental sciences both require light sources with shorter emission wavelengths: the former for improved resolution in photolithography and the latter for sensors that can detect minute hazardous particles. In addition, ultraviolet solid-state light sources are also attracting attention for potential applications in high-density optical data storage, biomedical research, water and air purification, and sterilization. Wide-bandgap materials, such as diamond² and III–V nitride semiconductors (GaN, AlGaN and AlN; refs 3–10), are potential materials for ultraviolet LEDs and laser diodes, but suffer from difficulties in controlling electrical conduction. Here we report the successful control of both n-type and p-type doping in aluminium nitride (AlN), which has a very wide direct band-gap¹¹ of 6 eV. This doping strategy allows us to develop an AlN PIN (p-type/intrinsic/n-type) homojunction LED with an emission wavelength of 210 nm, which is the shortest reported to date for any kind of LED. The emission is attributed to an exciton transition, and represents an important step towards achieving exciton-related light-emitting devices as well as replacing gas light sources with solid-state light sources.

Recently, we have achieved n-type conduction in Si-doped AlN (refs 12, 13). In the present work, by reducing the dislocation density and finely controlling the Si doping level, we were able to boost the room-temperature electron mobility. We examined (by Hall-effect measurement) the temperature dependence of the electron concentration (n) and electron mobility (μ_n) for n-type Si-doped AlN with a Si doping concentration of $3.5 \times 10^{17} \text{ cm}^{-3}$. At 300 K, the electron concentration was $7.3 \times 10^{14} \text{ cm}^{-3}$. The electron mobility was $426 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, which is the highest reported to date in n-type AlN and attests to the high quality of the AlN. As the temperature increased, the electron concentration increased exponentially and saturated at higher temperatures (Fig. 1a). The temperature dependence of electron concentration was fitted by the least-squares method, assuming the charge neutrality equation for n-type semiconductor with a shallow donor and a deep compensating acceptor¹². The best-fit values are donor concentration $N_D = 3.0 \times 10^{17} \text{ cm}^{-3}$, acceptor concentration $N_A = 2.6 \times 10^{16} \text{ cm}^{-3}$ and donor ionization energy $E_D = 282 \text{ meV}$. The donor concentration agreed well with the Si doping concentration, indicating that almost all Si atoms act as donors in AlN. The compensation ratio N_A/N_D was about 0.1. On the other hand, the electron mobility increased monotonically as temperature decreased (Fig. 1b). At 220 K, the electron mobility reached $730 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. In calculating mobility, we took into account scattering by neutral impurities, ionized impurities, polar optical phonons, acoustic deformation potential, and piezoelectric potential¹³. We assumed Matthiessen's rule, and used the best-fit

values of N_D , N_A and E_D , and the material parameters in ref. 13. The calculated electron mobility (solid line) agreed very well with the measured one. For n-type AlN with a higher dislocation density, the measured electron mobility was much smaller than the calculated one owing to the influence of scattering caused by charged dislocations¹⁴ or unidentified dislocation-related defects. Thus, the agreement indicates that the quality of the n-type AlN layer is high.

With this success in achieving n-type conduction, we turned our attention to p-type conduction and found that Mg doping produced p-type AlN. The p-type conduction in Mg-doped AlN was observed at Mg doping concentrations below $2 \times 10^{20} \text{ cm}^{-3}$. When the Mg doping concentration exceeded an upper limit of $2 \times 10^{20} \text{ cm}^{-3}$, the Mg-doped AlN became highly resistive. This is similar to the tendency observed in n-type Si-doped AlN (ref. 12), and probably results from a self-compensation effect of Mg in p-type AlN. This indicates that control of the Mg doping concentration is important for obtaining p-type conduction. In addition, the as-grown Mg-doped AlN was highly resistive, and became p-type conductive after thermal annealing in N_2 at 800 °C for 10 min. For the highly resistive as-grown Mg-doped AlN, the H concentration was in the same range as the Mg concentration, and after annealing the H concentration decreased. This means that, in the as-grown AlN, H passivated Mg dopant, but after annealing H was desorbed and Mg was activated, as in GaN (ref. 15).

We measured the temperature dependence of hole concentration (p) and hole mobility (μ_p) for Mg-doped AlN with a Mg doping concentration of $2 \times 10^{19} \text{ cm}^{-3}$. We clearly confirmed p-type conduction from the polarity of the Hall voltage, although there is some scattering of measured values because of the high resistivity (low

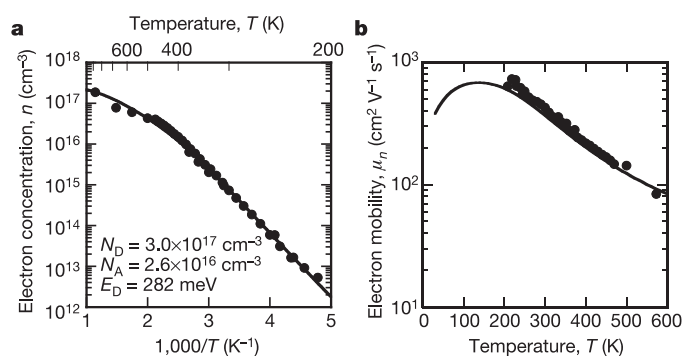


Figure 1 | Electrical characteristics of n-type Si-doped AlN. **a**, The temperature dependence of electron concentration. The solid line shows the least-squares fit, assuming the charge neutrality equation with a shallow donor and a deep compensating acceptor. The best-fit parameters are shown. **b**, The temperature dependence of electron mobility. The solid line shows the calculated mobility, considering specific scattering mechanisms (see text).

¹NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, 243-0198, Japan.

mobility). As the measurement temperature increased, the hole concentration increased exponentially (Fig. 2a). Accurately determining the acceptor concentration and compensating donor concentration in the p-type AlN would require fitting of the measured data in the high-temperature (exhausted) regime. Thus, from the present data, only the acceptor ionization energy E_A was obtained, by assuming that the hole concentration follows $\exp(-E_A/k_B T)$, where k_B is the Boltzmann constant and T is the sample temperature. The E_A was estimated to be 630 meV, which is in good agreement with an optically obtained value (510 meV)¹⁶ and with a theoretically obtained value (465–758 meV)¹⁷. Owing to the high acceptor ionization energy, the room-temperature hole concentration is as low as of the order of 10^{10} cm^{-3} . On the other hand, the hole mobility decreased as temperature increased (Fig. 2b). The temperature dependence was well fitted to $\mu \propto T^{-1.5}$ (broken line). This is typical of phonon scattering, which has the strongest influence on mobility in semiconductors at high temperature. To calculate the hole mobility, we used a method similar to the above-mentioned one for the electron mobility. In the calculation, we used the hole effective mass of AlN, $m_h^* = 3.3$, which we estimated from the acceptor ionization energy $E_A = 630 \text{ meV}$ and the effective mass theory. We assumed the Mg doping concentration of $2 \times 10^{19} \text{ cm}^{-3}$ as the acceptor concentration, and the compensation ratio of p-type AlN, $N_D/N_A = 0.1$. This ratio was the same as that of the n-type AlN. Although the experimental mobility was slightly smaller than the calculated one (solid line), both showed a similar tendency. These results also support p-type conduction in the Mg-doped AlN.

The achievement of n-type and p-type AlN allowed us to make a p-type AlN/intrinsic (undoped) AlN/n-type PIN AlN LED (Fig. 3a). The LED consists of an AlN PIN homojunction sandwiched between a p-type and an n-type AlN/AlGaIn superlattice¹⁸. The superlattices were used to improve the lateral conduction in the n-type region and reduce the contact resistance of the electrodes, and thereby lower the driving voltage of the LED. The average carrier concentration of the superlattices is as high as 10^{18} cm^{-3} , because the doping efficiency of the superlattices is much higher than that of the AlN layers. The high doping efficiency originates both from the large band offset between AlN and AlGaIn layers and from large band bending caused by strong polarization fields in the superlattices. A metal/undoped AlN (insulator)/n-type AlN (semiconductor) MIS LED was also fabricated to clarify the conduction and emission mechanisms of the LEDs (Fig. 3b).

The characterization of the electrical and optical properties was

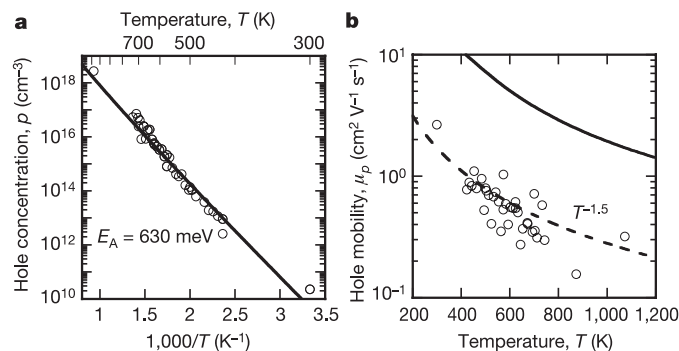


Figure 2 | Electrical characteristics of p-type Mg-doped AlN. **a**, The temperature dependence of hole concentration. The solid line shows the least-squares fit by $\exp(-E_A/k_B T)$, where E_A is the acceptor ionization energy, k_B the Boltzmann constant, and T the sample temperature. The best-fit parameter is shown. **b**, The temperature dependence of hole mobility. The broken line is a least-squares fit of data with $\mu \propto T^{-1.5}$. The solid line shows the calculated mobility, assuming the hole effective mass m_h^* of 3.3, the acceptor concentration N_A of $2 \times 10^{19} \text{ cm}^{-3}$, and the compensation ratio N_D/N_A of 0.1.

performed on-wafer under a direct current (d.c.) bias condition at room temperature. The current–voltage (I – V) characteristics of both LEDs showed rectifying properties, with rectification ratios of higher than 10^4 at $\pm 40 \text{ V}$ (Fig. 3c). The current increased remarkably above voltages V_D of 5 V for the PIN LED and 3.8 V for the MIS LED. These voltages are close to the diffusion (built-in) potential of the PIN and MIS junctions. Below V_D , the current was as low as $< 0.1 \mu\text{A}$. This low current is probably due to lower energy barriers at the metal/AlN interface and the n-type AlN/n-type superlattice interface. Above V_D , the current is limited by a high series resistance, which originates from the p-type and undoped AlN layers. Consequently, the operating voltages at 20 mA were about 45 V and 37 V for the PIN and MIS LEDs, respectively, and were much higher than the voltages estimated from the diffusion potentials. The operating voltage of the PIN LED is higher than that of the MIS LED because of a voltage drop in the p-type AlN as well as that in the undoped AlN.

For the PIN LED, the electroluminescence spectra were observed at a wavelength of approximately 210 nm (Fig. 3d). This emission wavelength was almost the same as that of the free-exciton recombination of an undoped AlN layer measured by photoluminescence at room temperature (Fig. 3e). We also assigned the excitonic structure of our AlN layer by cathodoluminescence and photoreflectance. The excitons in AlN are stable even at room temperature ($k_B T = 26 \text{ meV}$) because of the large exciton binding energy of 80 meV (ref. 11). This indicates that the origin of the electroluminescence is unambiguously

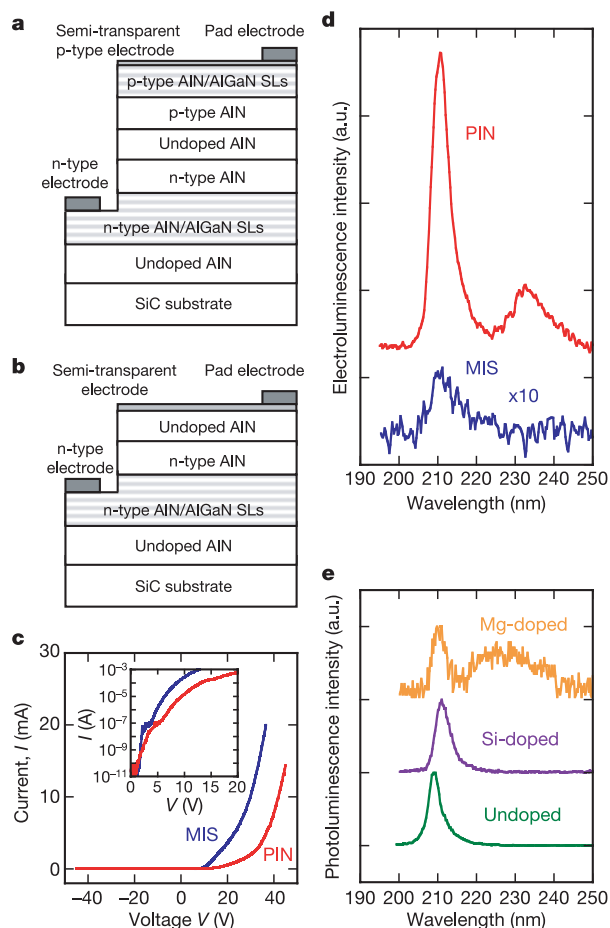


Figure 3 | AlN LEDs. **a**, Schematic illustration of the PIN LED. **b**, Schematic illustration of the MIS LED. **c**, Current–voltage (I – V) characteristics of the LEDs under a d.c. bias condition. The inset is plotted on a semilogarithmic scale. **d**, Electroluminescence spectra of the LEDs under a d.c. bias condition. **e**, Photoluminescence spectra of undoped, Si-doped and Mg-doped AlN. Data in **c**–**e** were measured at room temperature. a.u., arbitrary units; see text for explanation of PIN and MIS.

the near-band-edge recombination in the AlN. The electroluminescence intensity increased with current. As the current increased from 10 mA to 40 mA, the emission wavelength slightly shifted from 209 nm to 210 nm. This is because of a self-heating effect, which results from the relatively high operating voltage. From the temperature dependence of the free-exciton recombination energy¹⁹, the increase of junction temperature ΔT was estimated to be about 60 °C at 40 mA. In addition to the near-band-edge emission, a weak emission attributed to the Mg impurity level in the p-type AlN layer was observed at around 230 nm. On the other hand, a broad emission peak attributed to the Si impurity level in n-type AlN was also observed at around 400 nm.

We now discuss the conduction and emission mechanisms in the LEDs. For the MIS LED, emission at 210 nm was also observed. This indicates that holes are injected from the electrode by tunnelling and then recombine radiatively with electrons injected from the n-type AlN. The emission intensity of the PIN LED is more than 70 times as high as that of the MIS LED. This confirms that, for the PIN LED, p-type AlN and p-type superlattices improve the emission efficiency, and that the most of holes are injected from the p-type layers. On the other hand, the series resistance, $R_S = dV/dI$, was 10^6 – $10^5 \Omega$ at 6–10 V. The R_S is almost the same as a resistance of $5 \times 10^5 \Omega$ for hole transport in the p-type AlN, which was calculated from the result of the Hall-effect measurement. These indicate that, at low voltages, the diffusion and tunnelling processes are dominant in the conduction mechanism.

The value of R_S decreased with voltage. At high voltages of around 40 V, R_S decreased to about 500 Ω and has a linear relation with the inverse of current for both PIN and MIS LEDs. Such characteristics are commonly observed in PIN diodes operated under a high-voltage condition at which drift current limits the conduction²⁰. In our PIN LED, because the hole concentration in the p-type AlN is still low, only a few of the injected electrons recombine radiatively with the injected holes around the undoped AlN, while most of the electrons injected from the n-type AlN overflow through the undoped AlN to the p-type AlN and then to the p-type electrode. The Mg-related emission in electroluminescence supports this electron overflow. Therefore, at high voltage, electron drift current becomes dominant in the conduction mechanism. As a result, R_S is determined by the injected electron transport in the undoped and p-type AlN, and becomes smaller than that determined by the hole transport in the p-type AlN.

The output power of the near-band-edge emission at wavelength $\lambda = 210$ nm was measured to be about 0.02 μ W at a d.c. current of 40 mA. The external quantum efficiency η_{ext} of our on-wafer device was estimated to be of the order of $10^{-6}\%$, which is still lower than that for typical packaged commercial visible LEDs ($\eta_{\text{ext}} \approx 10\%$). The external quantum efficiency is given by the product of the extraction efficiency η_{out} and the internal quantum efficiency η_{in} , that is, $\eta_{\text{ext}} = \eta_{\text{out}} \times \eta_{\text{in}}$. The estimated η_{out} and η_{in} are of the order of $10^{-1}\%$ and $10^{-3}\%$, respectively. The reason for the low η_{out} is that emitted light is absorbed by the SiC substrate, the AlN/AlGaIn superlattices, and the p-type metal contact in the device. The η_{out} would be increased by using transparent substrates and proper packaging, and by optimizing the device structure.

However, increasing internal quantum efficiency η_{in} is much more important at this stage. Possible approaches to improving η_{in} effectively are to decrease dislocation density, introduce carrier confinement structures, and increase the p-type doping efficiency; we now consider these options in turn. First, our device still has a high dislocation density of 10^9 cm^{-2} as a result of the lattice mismatch and the heterovalency between the SiC substrate and the AlN layer. For AlN layers, in a dislocation-density range of 10^9 – 10^{10} cm^{-2} , the photoluminescence intensities, which are related to η_{in} , linearly increased with decreasing dislocation density. Thus, the dislocations act as non-radiative recombination centres. It has also been reported for LEDs made with other

nitride semiconductors that the output power increased by a factor of about 10^3 with decreasing dislocation density from 10^{11} to 10^7 cm^{-2} (ref. 21). The use of AlN substrate²², which has a low dislocation density of 10^4 cm^{-2} , will decrease the dislocation density in the device and thereby increase the η_{in} by a factor of 10^2 . Second, η_{in} generally increases by about 10^2 times by introducing carrier confinement structures (double heterostructures or quantum well structures), which are normally used for practical devices. BN in the zinc blende structure is an indirect bandgap material. The indirect (Γ –X transition) and direct (Γ – Γ transition) bandgap energies are 6.4 eV and about 8 eV, respectively²³, which are wider than AlN's direct bandgap. Therefore, BN or AlBN can confine carriers in the AlN active layer. Third, increasing the hole concentration in p-type AlN will enhance η_{in} because the hole concentration injected into the AlN active layer will increase. We have used Mg for p-type dopant, which is commonly used for nitride semiconductors, but Mg's doping efficiency in AlN is quite low. Alternative elements for the acceptor with lower ionization energy need to be explored—group II or IV elements (Be, Zn, Cd, or C) are potential candidates.

METHODS

Device manufacture. The layer structure of the AlN PIN LED (Fig. 3a) was grown on SiC(0001) substrate by low-pressure metal organic vapour phase epitaxy. The sources were trimethylaluminium, trimethylgallium and ammonia (NH_3). For n-type and p-type doping, Si and Mg atoms were doped by introducing silane (SiH_4) and bis-cyclopentadienylmagnesium (Cp_2Mg) during growth, respectively. We decreased the dislocation density to $2 \times 10^9 \text{ cm}^{-2}$ by suppressing the parasitic reaction of the sources (trimethylaluminium and NH_3) in the gas phase during the growth. The details of the growth procedure have been reported elsewhere^{12,13}. The PIN LED consisted of a 750-nm-thick undoped AlN layer, n-type uniformly Si-doped 100-period AlN/AlGaIn (1.2 nm/3 nm) superlattices with a Si doping concentration [Si] of $3 \times 10^{19} \text{ cm}^{-3}$, a 200-nm-thick n-type Si-doped AlN layer with [Si] of $2 \times 10^{18} \text{ cm}^{-3}$, a 60-nm-thick undoped (insulating) AlN layer, a 20-nm-thick p-type Mg-doped AlN layer with a Mg doping concentration [Mg] of $4 \times 10^{19} \text{ cm}^{-3}$, and p-type uniformly Mg-doped three-period AlN/AlGaIn (1.2 nm/3 nm) superlattices with [Mg] of $4 \times 10^{19} \text{ cm}^{-3}$. The structure of the MIS LED (Fig. 3b) is the same as that of the PIN LED except for the p-type AlN and p-type superlattices. A mesa structure $200 \mu\text{m} \times 200 \mu\text{m}$ in size was formed by dry etching with Cl_2 gas down to the n-type superlattices. An n-type Ti/Al/Ti/Au (25/100/50/100 nm) electrode was formed on the n-type superlattices, while a semi-transparent (transparency $T \approx 50\%$ at $\lambda = 210$ nm) p-type Pd/Au (2.5/1.5 nm) electrode was formed on the p-type superlattices for the PIN LED and on the undoped AlN for the MIS LED. A pad electrode was then evaporated on the Pd/Au electrode. **Measurements.** The mobility and carrier concentration were obtained by Hall-effect measurement in the Van der Pauw configuration. For the Hall-effect measurement of highly resistive (low mobility) samples, we applied an a.c. magnetic field to improve the signal-to-noise ratio. The electroluminescence spectra of the LED at various d.c. currents were measured with a monochromator equipped with a ultraviolet-enhanced charge coupled device array. The output power of the electroluminescence was measured with a Si photodiode located above the device. The output power of the near-band-edge emission was obtained by taking the intensity ratio between the near-band-edge emission and impurity-related emission into account.

Received 8 November 2005; accepted 24 March 2006.

- Schubert, E. F. & Kim, J. K. Solid-state light sources getting smart. *Science* **308**, 1274–1278 (2005).
- Koizumi, S., Watanabe, K., Hasegawa, M. & Kanda, H. Ultraviolet emission from a diamond p-n junction. *Science* **292**, 1899–1901 (2001).
- Akasaki, I. & Amano, H. Crystal growth and conductivity control of group III nitride semiconductors and their application to short wavelength light emitters. *Jpn. J. Appl. Phys.* **36**, 5393–5408 (1997).
- Pankove, J. I., Miller, E. A., Richman, D. & Berkeyheiser, J. E. Electroluminescence in GaN. *J. Lumin.* **4**, 63–66 (1971).
- Nakamura, S. et al. Superbright green InGaIn SQW structure LEDs. *Jpn. J. Appl. Phys.* **34**, L1332–L1335 (1995).
- Ponce, F. A. & Bour, D. P. Nitride-based semiconductors for blue and green light-emitting devices. *Nature* **386**, 351–359 (1997).
- Han, J. et al. AlGaIn/GaN quantum well ultraviolet light emitting diodes. *Appl. Phys. Lett.* **73**, 1688–1690 (1998).

8. Nishida, T. & Kobayashi, N. 346 nm emission from AlGa_N multi-quantum-well light emitting diode. *Phys. Status Solidi A* **176**, 45–48 (1999).
9. Adivarahan, V. *et al.* 250 nm AlGa_N light-emitting diodes. *Appl. Phys. Lett.* **85**, 2175–2177 (2004).
10. Hirayama, H. Quaternary InAlGa_N-based high efficiency ultraviolet light-emitting diodes. *J. Appl. Phys.* **97**, 091101 (2005).
11. Li, J. *et al.* Band structure and fundamental optical transitions in wurtzite AlN. *Appl. Phys. Lett.* **83**, 5163–5165 (2003).
12. Taniyasu, Y., Kasu, M. & Kobayashi, N. Intentional control of n-type conduction for Si-doped AlN and Al_xGa_{1-x}N (0.42 ≤ x < 1). *Appl. Phys. Lett.* **81**, 1255–1257 (2002).
13. Taniyasu, Y., Kasu, M. & Makimoto, T. Electrical conduction properties of n-type Si-doped AlN with high electron mobility (>100 cm² V⁻¹ s⁻¹). *Appl. Phys. Lett.* **85**, 4672–4674 (2004).
14. Ng, H. M. *et al.* The role of dislocation scattering in n-type GaN films. *Appl. Phys. Lett.* **73**, 821–823 (1998).
15. Nakamura, S., Mukai, T., Senoh, M. & Iwasa, N. Thermal annealing effects on p-type Mg-doped GaN films. *Jpn. J. Appl. Phys.* **31**, L139–L142 (1992).
16. Nam, K. B. *et al.* Mg acceptor level in AlN probed by deep ultraviolet photoluminescence. *Appl. Phys. Lett.* **83**, 878–880 (2003).
17. Mireles, F. & Ulloa, S. E. Acceptor binding energies in GaN and AlN. *Phys. Rev. B* **58**, 3879–3887 (1998).
18. Taniyasu, Y., Kasu, M., Kumakura, K., Makimoto, T. & Kobayashi, N. High electron concentrations in Si-doped AlN/AlGa_N superlattices with high average Al content of 80%. *Phys. Status Solidi A* **200**, 40–43 (2003).
19. Kuokstis, E. *et al.* Near-band-edge photoluminescence of wurtzite-type AlN. *Appl. Phys. Lett.* **81**, 2755–2757 (2002).
20. Sze, S. M. *Physics of Semiconductor Devices* 2nd edn 63–132 (Wiley, New York, 1981).
21. Amano, H. *et al.* Defect and stress control of AlGa_N for fabrication of high performance UV light emitters. *Phys. Status Solidi A* **201**, 2679–2685 (2001).
22. Rojo, J. C. *et al.* Report on the growth of bulk aluminium nitride and subsequent substrate preparation. *J. Cryst. Growth* **231**, 317–321 (2001).
23. Edgar, J. H. *Properties of Group III Nitrides* 129–134 (INSPEC, London, 1994).

Acknowledgements We thank M. Prinz, K. Thonke and R. Sauer at the University of Ulm in Germany for measurement of free-exciton energy in our AlN by cathodoluminescence and photoreflectance. We are also grateful to K. Kumakura for discussions, and K. Torimitsu and H. Takayanagi for their encouragement.

Author Contributions Y.T. performed experimental work and data analysis. All authors discussed results.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.T. (taniyasu@will.brl.ntt.co.jp).

Links between annual, Milankovitch and continuum temperature variability

Peter Huybers¹ & William Curry¹

Climate variability exists at all timescales—and climatic processes are intimately coupled, so that understanding variability at any one timescale requires some understanding of the whole. Records of the Earth's surface temperature illustrate this interdependence, having a continuum of variability following a power-law scaling^{1–7}. But although specific modes of interannual variability are relatively well understood^{8,9}, the general controls on continuum variability are uncertain and usually described as purely stochastic processes^{10–13}. Here we show that power-law relationships of surface temperature variability scale with annual and Milankovitch-period (23,000- and 41,000-year) cycles. The annual cycle corresponds to scaling at monthly to decadal periods, while millennial and longer periods are tied to the Milankovitch cycles. Thus the annual, Milankovitch and continuum temperature variability together represent the response to deterministic insolation forcing. The identification of a deterministic control on the continuum provides insight into the mechanisms governing interannual and longer-period climate variability.

Relative to the annual and Milankovitch cycles^{14,15}, the continuum of climate variability has received little attention, but is nonetheless expected to embody a rich set of physical processes¹². We focus on the continuum of surface temperature variability because of its pertinence to society and because of the relative abundance of records. The continuum can be described using spectral power laws: $P(f) \propto f^{-\beta}$, where P is spectral energy, f is frequency, and β is the power law. Power laws arise in physical systems as a result of nonlinearities transferring energy between different modes.

At interannual and decadal timescales, surface temperatures exhibit a strong land–sea contrast, with β averaging one over the oceans and zero over the continents^{3–5}. At millennial and longer timescales, ice-core and marine palaeoclimate records indicate β is closer to two^{1,3,6,7,12}. Thus, long-period temperature variability has a larger β than does short-period variability, with a transition occurring near centennial timescales. Note that ice-core palaeodata resolve timescales both shorter and longer than centennial periods, and demonstrate a change in the power law near centennial periods^{2,3}.

Numerous explanations have been advanced regarding the continuum, including stochastic resonance¹¹, modified random walks¹², and diffusion¹³, but all these models are partial descriptions of the variability. Ultimately, one seeks a physics of climate from which the temperature spectrum can be deduced, something analogous to oceanography's theories for tides¹⁶ and internal waves¹⁷. Further analysis of the continuum is expected to aid in distinguishing the processes that control climate change.

We use Thomson's multitaper method¹⁸ to estimate the surface temperature spectra (see Methods). Surface air temperatures (values at 2 m above the surface) from the NCEP-NCAR instrumental reanalysis¹⁹ are analysed from 1948 to 2002 and show, as expected, that β is smaller over land^{4,5}. Not previously documented is that β is also smaller toward high latitudes. For comparison, the average

land–sea contrast in β is 0.2, while the average Equator-to-pole difference is one. The average standard error in individual estimates of β is ± 0.16 .

The spatial patterns associated with β and the annual cycle mirror one another (Fig. 1). Zonally, the annual amplitude is smallest at the Equator, steeply increases to mid-latitudes, and then more gradually increases toward the poles; β has the opposite pattern. The Southern Hemisphere has a smaller annual amplitude and larger β than the Northern Hemisphere, presumably because of the greater ocean area. Similarly, the greater marine influence on the western margins of North America and Eurasia leads to a smaller annual amplitude and larger β than in the east.

Tropical ocean surface temperatures generally have a large β and a small annual cycle. That the largest values of β occur in the eastern equatorial Pacific is probably due to the El Niño variability. Note that El Niño does not obey a power law, but rather increases estimates of β by superimposing energy on the continuum at two- to five-year periods⁸. The presence of El Niño variability is also why the largest standard errors in β , ± 0.3 , occur in the eastern equatorial Pacific. The relatively large annual cycles in the eastern equatorial Pacific and Atlantic are primarily due to seasonal changes in upwelling²⁰.

The close relationship between the spectral energy at annual periods and the continuum is also evident when spectra are binned and averaged according only to annual period energy (Fig. 1e). Bins are designated from minimum to maximum annual energy at intervals of $0.2 \log_{10}[\text{°C}^2 \text{ds}^{-1}]$ (where ds is spectral bandwidth; see Methods) and, on average, contain 780 individual spectra. β decreases with the annual period energy, P_1 , as $\beta = -0.26 \log_{10}(P_1) + 0.81$. The cross-correlation between β and $\log_{10}(P_1)$ is -0.98 . Although β decreases with increasing P_1 , the average spectral energy between periods of 1.1 and 30 years, $\bar{P}$, actually increases. For spectral estimates with $P_1 > 1$, interannual and annual period energy are related as $\log_{10}(\bar{P}) = 0.58 \log_{10}(P_1) + 2.05$ with a cross-correlation of 0.99. Thus the three primary features of the instrumental temperature spectra (the power law, annual period energy, and interannual energy) are all tightly linked. Analysis of an instrumental compilation of surface temperatures²¹ yields similar relationships.

Further indication of a link between the annual cycle and the continuum comes from auto-bicoherence in surface temperature variability. Nonlinear interactions often produce phase dependencies between different frequencies. Auto-bicoherence measures nonlinear interaction by comparing the phase of two frequencies with the phase of their sum and difference frequencies²². An estimate of the global auto-bicoherence of surface air temperatures is obtained by averaging 330 individual estimates made using those records in an instrumental compilation²¹ having over 99% data coverage during the last century. A highly significant auto-bicoherence ($p = 0.01$) is found between the annual period and nearly all longer periods (Fig. 1f), supporting the inference that the temperature continuum is closely linked to the annual cycle. This bicoherence pattern can be

¹Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

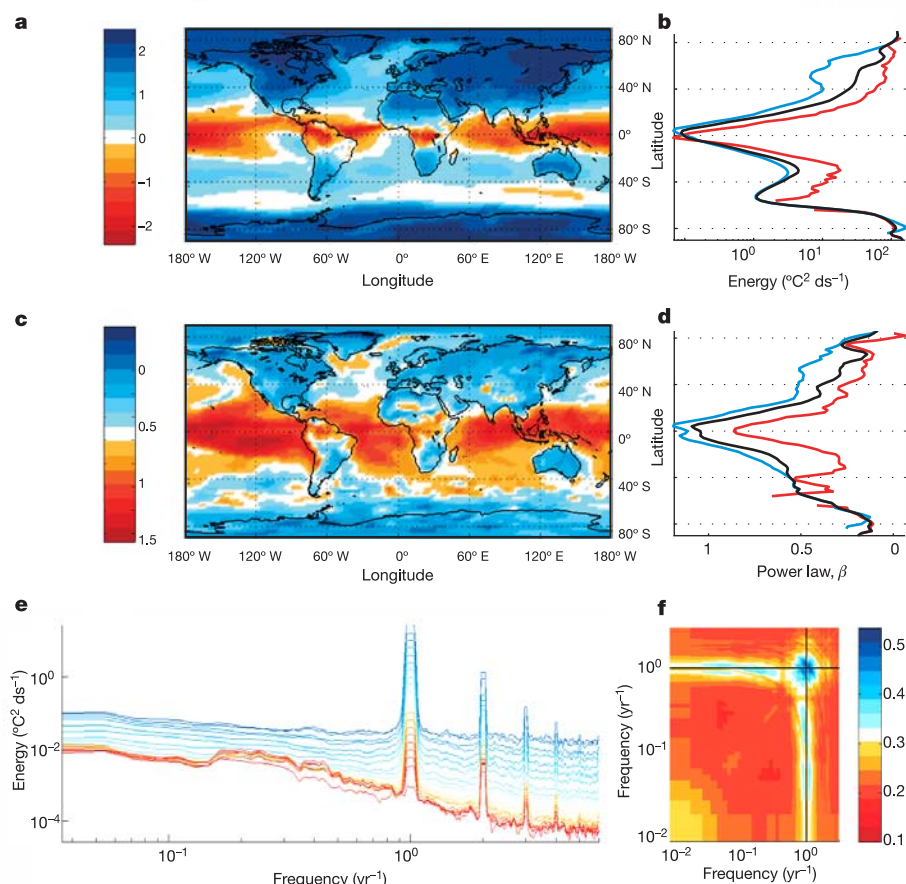


Figure 1 | Temperature scaling at instrumental periods from NCEP. **a**, Map of the spectral energy at the annual period in $\log_{10}$ units, and **b**, the zonal average (black) along with land-only (red) and ocean-only (blue) averages. **c**, β computed between two months and 30 years after removing the annual-period energy and its higher harmonics, and **d**, the associated zonal averages. Note that the colour scale in **c** and x axis in **d** are reversed.

e, Spectra binned according to annual-period energy and averaged. Line colours correspond to the colour scale in **a**. The β and average magnitude of the continuum both scale with the amplitude of the annual cycle. **f**, Auto-bicoherence estimate of instrumental surface air temperatures²¹. The approximate 99% confidence level is 0.3. Black lines indicate the annual cycle.

modelled using multiplicative noise. For example, if weather variations are modelled as a stochastic white noise process and then multiplied by an annual cycle such that variability is greater during the winter, an auto-bicoherence pattern similar to Fig. 1f is obtained.

The two regions having the greatest contrast in temperature spectra are the tropical oceans and the high-latitude continents (Fig. 1). To trace the behaviour of these contrasting environments out to longer timescales, two separate sets of palaeorecords are compiled, one from each region, both spanning monthly to million-year timescales (see Fig. 2 and Supplementary Information). Although noise and other climatic effects are present, the simplest interpretation of each palaeorecord is as a proxy for temperature. Spectral estimates at most frequency bands are available from at least two independent proxies and agree well with one another, providing confidence in the result. We first discuss the concentrations of spectral energy (that is, the peaks) in the proxy temperature spectrum and then the continuum.

Both the high-latitude and tropical spectra show significant ($p = 0.05$) spectral peaks associated with insolation forcing. At high latitudes, the annual cycle is massive, accounting for 80% of the variance at monthly to million-year timescales, while in the tropics the annual cycle accounts for only 25% of the total variance. Periods longer than 10 kyr contain 10% of the high-latitude variance and 50% of the tropical variance. At long periods, a 23-kyr precession period is found in tropical planktonic $\delta^{18}\text{O}$ records and, to a lesser extent, Mg/Ca records. 41-kyr obliquity variations are directly present in the insolation forcing, and correspond to peaks in tropical

planktonic $\delta^{18}\text{O}$, tropical alkenone, and high-latitude ice-core $\delta^{18}\text{O}$ records. The most energetic low-frequency band is associated with the ~ 100 -kyr glacial cycles.

Turning now to the continuum of temperature variability, between annual and centennial periods the tropical marine compilation has $\beta = 0.56 \pm 0.08$ while the high-latitude compilation has $\beta = 0.37 \pm 0.05$, both in agreement with results shown in Fig. 1. The magnitudes of the tropical and high-latitude spectra approach one another on moving from annual to longer periods and nearly converge at centennial periods. At periods longer than centuries, however, steeper power-law relationships exist. In the tropics $\beta = 1.29 \pm 0.13$ and at high latitudes $\beta = 1.64 \pm 0.04$. With this change in slope the spectra diverge as one moves from centennial to longer timescales. The tropical and high-latitude continua are most similar at centennial timescales. Conversely, the high-latitude continuum is more than an order of magnitude more energetic near the annual and Milankovitch periods. Apparently, the continuum is more energetic at locations that have greater insolation variability and at timescales neighbouring those of the insolation forcing.

Whereas the annual and Milankovitch-period climate variations tell us primarily about the forcing, the continuum tells us about the dynamical processes governing climate variability. The existence of two separate scaling regimes above and below centennial periods suggests the presence of distinct controls on the climate variability. Possibly, the climate system has a memory associated with the oceans that causes high-frequency variability to accumulate into progressively larger and longer-period variations²³, and a Milankovitch-driven low-frequency

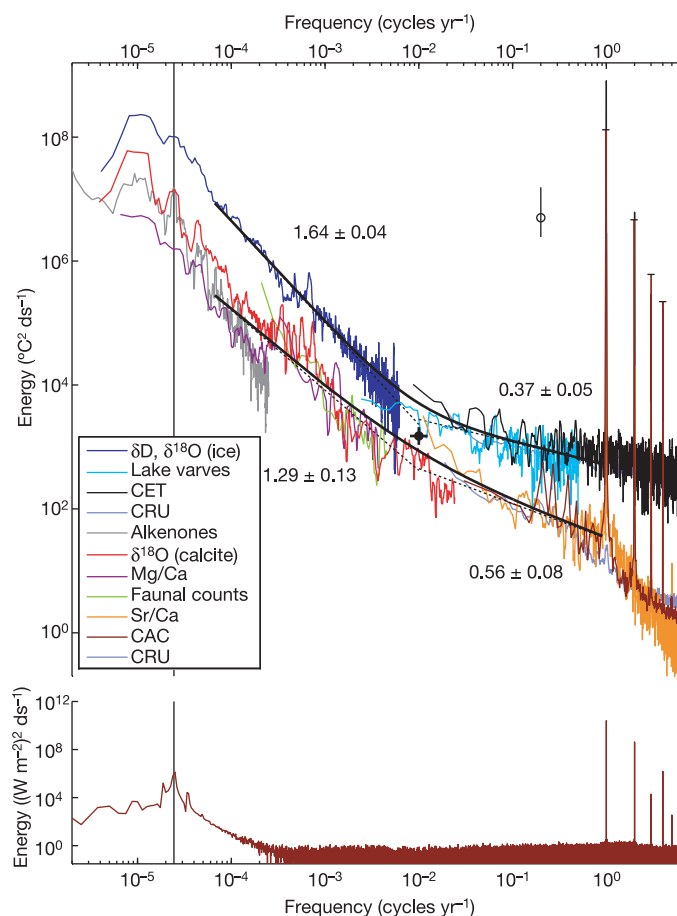


Figure 2 | Patch-work spectral estimate using instrumental and proxy records of surface temperature variability, and insolation at 65° N. The more-energetic spectral estimate is from high-latitude continental records and the less-energetic estimate from tropical sea surface temperatures. High-latitude spectra are estimated from Byrd, Taylor and GISP2 ice-core $\delta^{18}\text{O}$; Vostok and Dome C ice-core δD ; Donard lake varve thickness; the Central England Time-series (CET); and the Climate Research Unit's (CRU) instrumental compilation. From low latitudes we use ODP846 marine sediment-core alkenones; W167-79, OCE205-103, EW9209-1, ODP677 and ODP927 calcite $\delta^{18}\text{O}$; PL07-39 and TR163-19 Mg/Ca; ODP658 foram assemblages; Rarotonga coral Sr/Ca; and the Climate Analysis Center (CAC) and CRU instrumental compilations. Temperature spectral estimates from records of the same data type are averaged together ($\delta^{18}\text{O}_{\text{ice}}$ and $\delta\text{D}_{\text{ice}}$ are also averaged). Power-law estimates between 1.1–100 yr and 100–15,000 yr periods are listed along with standard errors, and indicated by the dashed lines. The sum of the power-laws fitted to the long- and short-period continuum are indicated by the black curve. The vertical line-segment indicates the approximate 95% confidence interval, where the circle indicates the background level. The mark at 1/100 yr indicates the region mid-way between the annual and Milankovitch periods. At bottom is the spectrum of insolation at 65° N sampled monthly over the past million years plus a small amount of white noise. The vertical black line indicates the 41-kyr obliquity period.

response that transfers spectral energy toward higher frequencies, possibly involving nonlinear ice-sheet dynamics. These low- and high-frequency temperature responses appear to be of nearly equal magnitude at centennial timescales, midway in log-frequency space between the annual and Milankovitch bands. Interannual temperature variability may also follow different power laws during glacial and interglacial periods, but assessment of this further possibility awaits a global set of high-resolution glacial climate records.

What would the background continuum look like in the absence of annual or Milankovitch variations? Variations would presumably still exist owing to adjustments in the Equator to pole temperature

gradient, providing a background of climate variability upon which variations due to Milankovitch- and annual-period changes are superimposed. The diurnal cycle is also expected to cause a local continuum response, but we have restricted the analysis to the more climatically relevant longer-period responses.

Discussion of long-term climate variability is commonly divided between deterministic¹⁵ and stochastic components¹²—often associated with spectral peaks and the continuum, respectively¹⁰. The analysis presented here indicates that the annual and Milankovitch energy are linked with the continuum, and together represent the climate response to insolation forcing.

These observations have implications for understanding future climate variability. The distinction between the magnitude of low- and high-latitude climate variability²⁴ may become smaller toward centennial timescales. Furthermore, the physics controlling centennial and longer-timescale temperature variability appears to be distinct from that of shorter timescales, raising questions of whether climate models able to represent decadal variations²⁵ will adequately represent the physics of centennial and longer-timescale climate variability. Finally, because the annual and Milankovitch cycles are deterministic, a greater understanding of their influence on the continuum may lead to improved predictability of interannual and longer-timescale climate variability.

METHODS

Spectral analysis. Spectra are estimated using Thomson's multitaper method¹⁸ with three windows. To reduce the influence of a small bias present at the lowest frequencies in the multitaper spectral estimates²⁶, the three lowest-frequency estimates are dropped when estimating β . For palaeorecords, when spectral estimates from the same proxy type overlap, a weighted average is taken. Weightings are proportional to frequency resolution so that better resolved records contribute more heavily. Spectral energy, or more precisely spectral energy density, is measured in $^{\circ}\text{C}^2 \text{ ds}^{-1}$. In this case energy refers to the squared quantity, $^{\circ}\text{C}^2$ (not joules), and density refers to energy per unit frequency, ds, where ds equals one over the length of the record in years. The energy density of the spectral continuum is independent of record duration.

The power law, β , is estimated by a least-squares fit to log-frequency and log-power-density estimates. To more uniformly weight the estimate, spectra are binned into equally spaced log-frequency intervals and averaged before fitting β . For the instrumental records, monthly anomalies are used when estimating β in order to filter out spectral energy associated with the annual cycle and its higher harmonics. To preclude effects from a small residual annual cycle, power-law fits and computation of the average continuum energy exclude frequencies between 1/0.8 to 1/1.2 yr^{-1} and 1/1.8 to 1/2.2 yr^{-1} . For palaeorecords, β is estimated between 1/100 and 1/15,000 yr^{-1} to minimize the influence from the Milankovitch bands. Uncertainties reported for β are the standard error of the least-squares fit. In the case of the auto-bicoherence statistic, the energy contained at each frequency band is first normalized to unity so that only the phase information is retained, thus ensuring that the magnitude of the annual cycle does not bias the result.

The fraction of variance at the annual cycle is calculated by integrating the estimated spectral energy between frequencies of 1/0.95 to 1/1.05 and then dividing by the total spectral energy. The integration accounts for variations in the frequency spacing so that changes in record length do not influence estimates of the energy at the annual period. At high frequencies, only instrumental estimates are used as these are more accurate. Furthermore, tropical $\delta^{18}\text{O}_{\text{calcite}}$ records are excluded to minimize influence from non-temperature effects. To permit direct comparison between the continuum and annual-period energy in Fig. 2, the amplitude of the annual cycle and its harmonics are increased to the expected value for a record one million years in duration.

Another technique commonly used to estimate scaling relationships is detrended fluctuation analysis^{4,5,7} (DFA), which yields a coefficient, α , related to the spectral power law by $\alpha = (\beta + 1)/2$. Results presented here were checked using DFA and are in agreement; this is as expected, because DFA and Fourier estimates are mathematically equivalent up to the choice of weighting terms²⁷. For ease of comparison, other authors' DFA results are reported as β values.

Data. The NCEP reanalysis¹⁹ of two-metre air temperatures has global coverage but only spans the period 1949 to 2002. Palaeo-temperature proxies are necessary to resolve centennial and longer-period temperature variability. Following ref. 1, we piece together spectral estimates from records of differing durations and resolution. Two separate spectral patchworks of surface air

temperature are made: one using high-latitude continental records and another using tropical marine records. At high latitudes, $\delta^{18}\text{O}_{\text{ice}}$ ice-core records and lake sediment data constrain surface air temperature variability. Tree-ring records are excluded because low-frequency variability can only be reconstructed over large spatial scales²⁸, and this analysis is restricted to localized temperatures. In the tropics, we use foram assemblages, Sr/Ca from corals, alkenones, $\delta^{18}\text{O}_{\text{calcite}}$ and Mg/Ca_{calcite} records to constrain sea surface temperature variability. References for the records and pertinent statistics are provided in Supplementary Information.

The entire $\delta^{18}\text{O}_{\text{calcite}}$ signal is converted into temperature variability, not correcting for the ice-volume component, because we expect higher-frequency variability to reflect rapid changes in temperature and no frequency-dependent correction is available. At low frequencies, this leads to an overestimate in the amplitude of $\delta^{18}\text{O}_{\text{calcite}}$ temperatures by roughly a factor of two, but at high frequencies $\delta^{18}\text{O}_{\text{calcite}}$ variability is in good agreement with the other temperature proxies. Spectral power laws are estimated in this higher-frequency region ($>1/15$ kyr), mitigating effects from the ice-volume component of the $\delta^{18}\text{O}_{\text{calcite}}$ variability.

There are numerous uncertainties associated with the interpretation of palaeo-proxies. Uncertainties directly influencing the shape of the spectra include rectification and aliasing²⁹—potentially first order effects, considering that most proxies are affected by but do not resolve the annual variability. Spectral estimates are also subject to errors or assumptions built into the age model³⁰. All records use their published age models except for ODP677 and ODP927 where a depth-derived age-model³⁰ is used, not making orbital assumptions.

The scope of error due to non-temperature effects, observational uncertainty, age-model uncertainty, aliasing and rectification is unclear. However, some confidence is provided in that various proxies, sensitive to differing environmental factors and measured at different locations, yield commensurate spectral estimates at most frequency bands. Furthermore, at annual to centennial timescales, the proxies reproduce the same scaling structure evident in the instrumental records.

Received 6 November 2005; accepted 23 March 2006.

- Shackleton, N. J. & Imbrie, J. The $\delta^{18}\text{O}$ spectrum of oceanic deep water over a five-decade band. *Clim. Change* **16**, 217–230 (1990).
- Ditlevsen, P., Svensmark, H. & Johnsen, S. Contrasting atmospheric and climate dynamics of the last-glacial and Holocene periods. *Nature* **379**, 810–812 (1996).
- Pelletier, J. The power-spectral density of atmospheric temperature from time scales of 10^{-2} to 10^6 yr. *Earth Planet. Sci. Lett.* **158**, 157–164 (1998).
- Fraedrich, K. & Blender, R. Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.* **90**, 108501 (2003).
- Blender, R. & Fraedrich, K. Long time memory in global warming simulations. *Geophys. Res. Lett.* **14**, doi:10.1029/2003GL017666 (2003).
- Schmitt, F., Lovejoy, S. & Schertzer, D. Multifractal analysis of the Greenland Ice-core Project climate data. *Geophys. Res. Lett.* **22**, 1689–1692 (1995).
- Ashkenazy, Y., Baker, D., Gildor, H. & Havlin, S. Nonlinearity and multifractality of climate change in the past 420,000 years. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL018099 (2003).
- Philander, G. *El Niño, La Niña, and the Southern Oscillation* (Academic, San Diego, California, 1990).
- Hurrell, J. & Van Loon, H. Decadal variations in climate associated with the North Atlantic Oscillation. *Clim. Change* **36**, 301–326 (1997).
- Kominz, M. & Pisias, N. Pleistocene climate—deterministic or stochastic? *Science* **204**, 171–173 (1979).
- Pelletier, J. Coherence resonance and ice ages. *J. Geophys. Res.* **108**, doi:10.1029/2002JD003120 (2003).
- Wunsch, C. The spectral description of climate change including the 100 ky energy. *Clim. Dyn.* **20**, 353–363 (2003).
- Fraedrich, K., Luksch, U. & Blender, R. A 1/f-model for long time memory of the ocean surface temperature. *Phys. Rev. E* **70**, 037301 (2003).
- Hays, J., Imbrie, J. & Shackleton, N. Variations in the earth's orbit: Pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Imbrie, J. *et al.* On the structure and origin of major glaciation cycles. 1. Linear responses to Milankovitch forcing. *Paleoceanography* **6**, 205–226 (1992).
- Wunsch, C. The spectrum from two years to two minutes of temperature fluctuations in the main thermocline at Bermuda. *Deep-Sea Res.* **19**, 577–593 (1972).
- Garrett, C. & Munk, W. Internal waves in the ocean. *Annu. Rev. Fluid Mech.* **80**, 291–297 (1979).
- Percival, D. & Walden, A. *Spectral Analysis for Physical Applications* (Cambridge Univ. Press, Cambridge, UK, 1993).
- Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- DeWitt, D. G. & Schneider, E. K. The processes determining the annual cycle of equatorial sea surface temperature: A coupled general circulation model perspective. *Mon. Weath. Rev.* **127**, 381–395 (1999).
- Jones, P., New, M., Parker, D., Martin, S. & Rigor, I. Surface air temperature and its variations over the last 150 years. *Rev. Geophys.* **37**, 173–199 (1999).
- Elgar, S. & Sebert, G. Statistics of bicoherence and biphasic. *J. Geophys. Res.* **94**, 10993–10998 (1989).
- Hasselmann, K. Stochastic climate models. Part I. Theory. *Tellus* **6**, 473–485 (1976).
- Serreze, M. C. *et al.* Observational evidence of recent change in the northern high-latitude environment. *Clim. Change* **46**, 159–207 (2000).
- Delworth, T. L. *et al.* Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *J. Clim.* **19**, 555–574 (2002).
- McCoy, E., Walden, A. & Percival, D. Multitaper spectral estimation of power law processes. *IEEE Trans. Signal Process.* **46**, 655–668 (1998).
- Wilson, K., Francis, D., Wensel, R., Coats, A. & Parker, K. Relationship between detrended fluctuation analysis and spectral analysis of heart-rate variability. *Physiol. Meas.* **23**, 385–401 (2002).
- Briffa, K. *et al.* Low-frequency temperature variations from a northern tree-ring density network. *J. Geophys. Res.* **106**, 2929–2941 (2001).
- Wunsch, C. & Gunn, D. A densely sampled core and climate variable aliasing. *Geo-mar. Lett.* **29**, doi:10.1007/s00367-003-0125-2 (2003).
- Huybers, P. & Wunsch, C. A depth-derived Pleistocene age-model: Uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography* **19**, doi:10.1029/2002PA000857 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements P.H. was supported by the NOAA Postdoctoral Program in Climate and Global Change. Funding for W.C. was provided by the National Science Foundation, Division of Ocean Sciences. T. Crowley, R. Ferrari, O. Marchal, J. Sachs, D. Steele and C. Wunsch provided useful comments.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.H. (phuybers@whoi.edu).

Seismic detection of folded, subducted lithosphere at the core–mantle boundary

Alexander R. Hutko¹, Thorne Lay¹, Edward J. Garnero² & Justin Revenaugh³

Seismic tomography has been used to infer that some descending slabs of oceanic lithosphere plunge deep into the Earth's lower mantle^{1,2}. The fate of these slabs has remained unresolved, but it has been postulated that their ultimate destination is the lowermost few hundred kilometres of the mantle, known as the D'' region. Relatively cold slab material may account for high seismic velocities imaged in D'' beneath areas of long-lived plate subduction, and for reflections from a seismic velocity discontinuity just above the anomalously high wave speed regions^{3,4}. The D'' discontinuity itself is probably the result of a phase change in relatively low-temperature magnesium silicate perovskite^{5,6}. Here, we present images of the D'' region beneath the Cocos plate using Kirchhoff migration of horizontally polarized shear waves, and find a 100-km vertical step occurring over less than 100 km laterally in an otherwise flat D'' shear velocity discontinuity. Folding and piling of a cold slab that has reached the core–mantle boundary, as observed in numerical and experimental models, can account for the step by a 100-km elevation of the post-perovskite phase boundary due to a 700 °C lateral temperature reduction in the folded slab. We detect localized low velocities at the edge of the slab material, which may result from upwellings caused by the slab laterally displacing a thin hot thermal boundary layer.

The Earth's D'' region contains a large thermal boundary layer caused by heat flowing from the molten iron–alloy outer core into the crystalline silicate/oxide mantle. However, a thermal boundary alone cannot account for all seismically detected structures in D'' (ref. 3), such as an abrupt increase in seismic velocities detected at heights varying from 130 to 340 km above the core–mantle boundary (CMB). Shear (S) and compressional (P) wave velocity increases at this D'' discontinuity range from 1–3% and 0.5–3%, respectively, and vary spatially. A recently discovered phase transition in (Mg,Fe)SiO₃ perovskite is predicted to occur at pressures and temperatures near the top of the D'' region, offering a viable explanation for the D'' discontinuity^{5,6}. It is important to assess whether thermal and phase-change effects can account for the discontinuity, or whether additional chemical heterogeneity is required^{3,7–9}.

P and S wave global tomography studies indicate an anomalously fast quasi-tabular structure extending deep into the lower mantle beneath the Caribbean^{1,2}, attributed to the thermal effect of remnant Farallon slab, which is comprised of former eastern Pacific basin oceanic lithosphere. If this material reaches the CMB, it should elevate the velocity discontinuity owing to the large positive Clapeyron slope of the perovskite/post-perovskite phase transition¹⁰. Seismological investigations of the D'' region below the localized region under the Cocos plate have found small-scale variations in D'' velocity structure using S and P waves^{11,12}, apparent topography of a pervasive D'' shear velocity reflector^{13,14}, and apparent anisotropy in D'' (ref. 15).

Given the uncertain fate of slab material, we employ an imaging approach that makes minimal assumptions about the geometry of structures that give rise to seismic wave arrivals, Kirchhoff migration, or diffraction-stacking, which allows seismic waves to originate anywhere within a three-dimensional volume. This method is ideal for imaging unknown deep mantle slab configurations^{13,16–19}, discontinuities with strong topography or limited lateral extent, and localized heterogeneities such as plumes rising from the CMB. We apply Kirchhoff migration to a data set previously processed for one-dimensional stacking¹¹ to determine the lowermost-mantle shear velocity structure beneath the Cocos Plate (Fig. 1). The data have stable horizontal shear (SH) radiation of S and core-reflected ScS waves, and are processed to obtain dominant periods of about 3 s.

Our Kirchhoff approach¹⁷ assumes isotropic single-scattering: when an incident SH wave encounters a point scatterer, a scattered

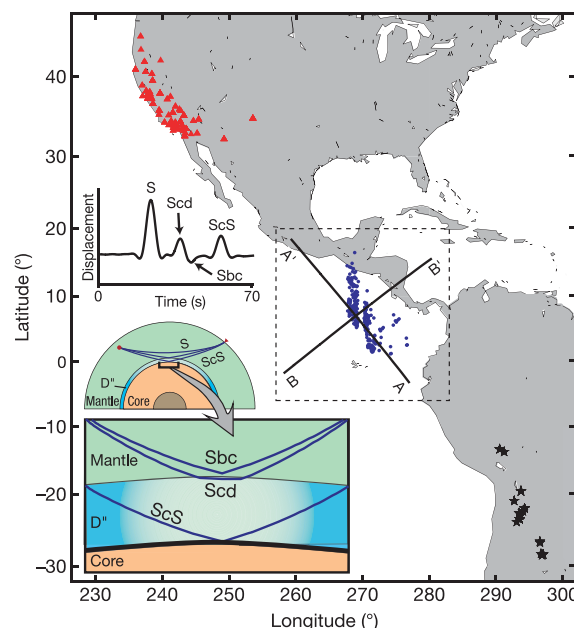


Figure 1 | Map showing earthquake epicentres (black stars), seismic station locations (red triangles) and ScS reflection points (blue dots). Also shown are the two lines along which cross-sections are made through the migration image volume in the lower mantle (black lines). The bottom left inset shows S and ScS raypaths through the mantle. At long epicentral distances, the direct S wave turns deep in the lowermost mantle and the phases interacting with the D'' discontinuity have their largest amplitudes. A representative synthetic seismogram is shown, illustrating the small negative overshoot of the Scd + Sbc arrival between S and ScS arrivals.

¹Earth Sciences Department, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. ²Department of Geological Sciences, Arizona State University, Box 871404, Tempe, Arizona 85287-1404, USA. ³Department of Geology and Geophysics, University of Minnesota, 310 Pillsbury Dr. SE, Minneapolis, Minnesota 55455, USA.

signal propagates to receivers without further interaction with other scatterers. This is valid for velocity heterogeneity strengths less than a few per cent. The travel times from every source and receiver are calculated to possible scattering locations within a three-dimensional grid emplaced in the lower mantle. The differential time between a given scattering location arrival and either S or ScS is computed for time-shifting each seismogram relative to the reference phase. Shifted seismograms for all source–receiver combinations are linearly summed to obtain an estimate of the wave amplitude associated with each candidate scattering location. We only consider energy that is scattered with upward-directed raypaths to the source and receiver, which excludes possible polarity reversals caused by underside reflections or scattering.

Vertical cross-sections (Fig. 2) through the data migration volume show strong reflectance associated with the CMB and the D'' discontinuity. The CMB is illuminated by ScS, and as expected, imaged as an extensive, nearly horizontal scattering surface. The D'' discontinuity is illuminated by large positive-amplitude post-critical refracted phases (called Scd, ref. 20), and small negative-amplitude phase-shifted post-critical reflected phases (Sbc, see Fig. 1). The D'' discontinuity is 160–190 km above the CMB towards the southeast,

and 250–290 km above the CMB in the northwest of our study region. We note that migrations directly depend on the assumed reference model: the discontinuity depth maps 50–100 km deeper using the PREM²¹ reference model than it does for migrations using models with higher velocities in D''. The apparent D'' discontinuity depth differs by up to 70 km depending on whether S or ScS is the reference phase. However, the abrupt step is unequivocally present in all cases and the step amplitude changes by less than 20 km. This stability adds confidence that the step is not an artefact of earthquake mislocation or localized regions of velocity heterogeneity sampled by the reference phase. Kirchhoff migration theory assumes pre-critical reflections and does not account for wavefield triplication effects from a velocity discontinuity within the image volume; it thus slightly underestimates depth and blurs the image of the reflector because the higher-amplitude, refracted Scd phase turns below the discontinuity, arriving slightly before reflected energy. Even so, these are the highest-resolution three-dimensional deep-mantle reflector images of the lower mantle so far, adding refined focus to past work that infers a larger sloping discontinuity gradient in this region¹³.

To address the trade-off between discontinuity depth and volumetric heterogeneity we compute travel-time perturbations through mantle tomography models (UT², UCB²², CIT²³) for the unperturbed raypaths of both reference phases and the scattering raypath geometries (Fig. 3). Images for different tomography models suggest lateral disruption of the discontinuity near 6° N, and vertical offsets across

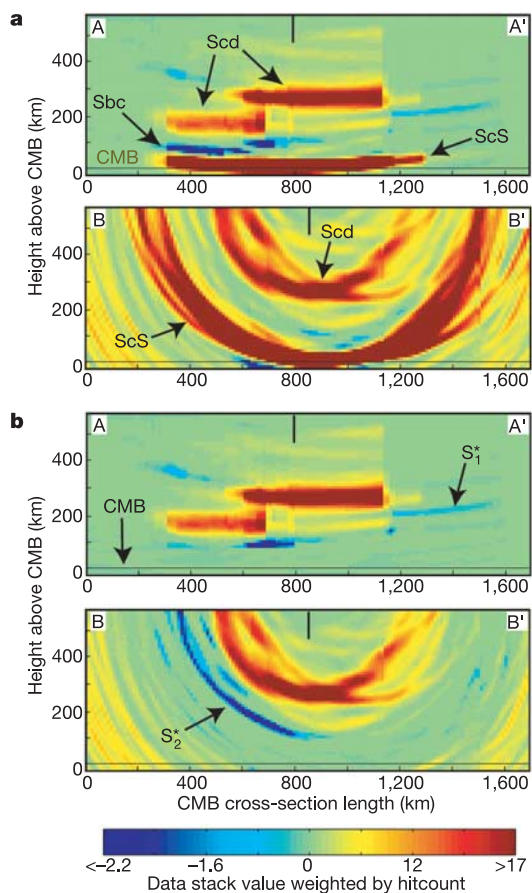


Figure 2 | Vertical cross-sections through the scattering image volume for our data set migrated in the PREM model using ScS as the reference phase. The horizontal axis in each panel is lateral distance going from A–A' and B–B' spanning approximately 1,700 km. The black tick mark indicates where the two profiles cross each other. The D'' discontinuity can be seen in **a** along A–A' at about 180 and 280 km above the CMB in the southeast and northwest parts of the image, respectively. The images in **b** are the same as in **a** except that the ScS energy has been censored by 4-s-wide tapers. All data stack values (the sum of seismogram amplitudes aligned along travel times corresponding to a scattered phase originating from a given grid point) have been weighted according to the hitcount (the number of contributing seismograms). There is no vertical exaggeration.

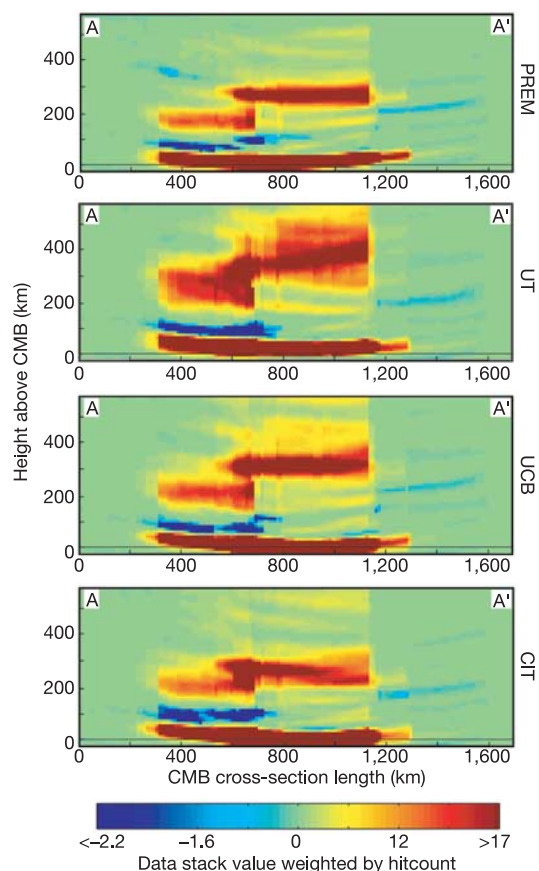


Figure 3 | Vertical cross-sections along profile A–A' through the image volume migrated relative to ScS using four different velocity models. Models are identified on the right axis. The direct S wave energy has been suppressed by 4-s-wide tapers. Data stack values have been weighted according to the hitcount. Profiles are chosen along the ray-path corridor (A–A'), because oblique profiles through the image volume will blend together the well-resolved in-plane images with more poorly resolved out-of-plane images, which have much more streaking.

the discontinuity step of 50 to 150 km. The discontinuity feature imaged using the UT model is spread out vertically and located at shallower depth than for the PREM migrations. These effects stem from relatively high lowermost-mantle shear velocities of the UT model. This should not be construed as a deterioration of the images relative to those found using PREM; lower reference velocities intrinsically sharpen reflector images. The CIT model, which has a northward decrease in velocity in the D'' layer, tends to suppress the vertical offset by about 50 km, but does not eliminate it. However, this model does not predict a northward decrease of ScS–S differential times, as is clearly observed in the data. ScS–S differential time predictions from the UT and UCB models are consistent with the data; the UT and UCB models increase the imaged discontinuity step height relative to the PREM-derived image by about 10 km.

Our migrations cannot preclude the existence of small-scale (<30 km vertically) undulations in the reflecting surface. However, images for all reference models (Fig. 3) are compatible with a quasi-horizontal reflector over a horizontal scale of 800 km offset by a step in the discontinuity depth. The multiple-event stacking here strongly favours an abrupt feature (occurring over less than 100 km laterally) rather than the smooth feature inferred in ref. 13 which averaged highly dispersed reflector images from small subsets of data for individual events.

Cross-sections through the scattering volume also show weak negative amplitude (blue) features with significant coherence. A fairly narrow blue streak appears below the discontinuity (red) in the southeast. This feature is concentrated in the plane of propagation and is a predictable artefact of the phase-shifted post-critical reflection associated with the overlying discontinuity, as confirmed by migrating synthetic seismograms. A small negative-amplitude feature (S_2^*) also extends northwestward from the northern end of the positive reflector at a height of about 220 km above the CMB. This feature is faint because only data from the two northernmost earthquakes adequately sample it. The image is consistent with streaking from a low-velocity point-scatterer on the margin of the illuminated zone. This feature is not present in the three global tomography models; however, a recent finite frequency regional tomography model²⁴ shows evidence for it. Other faint positive- and negative-amplitude features (Fig. 2), especially shallow ones, may be crustal multiples that persist due to incomplete suppression of receiver coda following the direct S wave.

A prominent negative-amplitude feature (S_2^*) is present in migration volume cross-sections perpendicular to the raypaths when ScS arrivals are muted out (Fig. 2b). These cross-sections display the classic ‘smiles’ of scattering migrations that appear when the data sample only a limited azimuthal corridor, as in our case. Although the location is not well resolved, the negative-amplitude feature is strongest at a height of about 200 km above the CMB, on the southwest edge of the northwestern, shallower positive discontinuity (see Supplementary Information for more cross-sections). Using a migration approach with smaller data sets, a prior study implied that a negative reflector exists below the positive reflector in the northwest¹³. Our results show very little coherent negative in-plane energy at depths below the D'' discontinuity towards the northwest, which is consistent with results from one-dimensional waveform stacking¹¹; however, all of our models indicate that a negative-amplitude scatterer is offset to the southwest. Out-of-plane scattering produces an arrival, readily apparent in waveforms from central California stations¹³ that approaches the ScS arrival and does not stack coherently for in-plane sections or one-dimensional stacks (see Supplementary Information).

The cause of the D'' discontinuity remains uncertain³. There is currently much enthusiasm for interpreting the P and S velocity discontinuities near the top of D'' as the result of the (Mg,Fe)SiO₃ perovskite to post-perovskite phase transformation^{5,6}. This phase change is predicted to have a large positive Clapeyron slope of about 7.5 MPa K⁻¹ (ref. 10). To account for a 100-km step in the D''

discontinuity by purely thermal effects, a lateral temperature gradient of 700 K is required over less than ~100 km laterally. Such a large lateral temperature gradient clearly requires a dynamic environment. Recent geodynamic calculations²⁵ indicate that slabs descending to the base of the mantle can retain strong lateral temperature gradients and buckle and spread out as they approach the CMB. Folding of the Farallon slab and spreading to the west offers a viable explanation for the strong lateral temperature gradients that interact with the phase-change sensitivity to produce the observed step in the D'' discontinuity (Fig. 4).

A recent study proposed a vertical pairing of overlying positive and negative velocity discontinuities due to transformation to post-perovskite at the top of D'' and then back to perovskite in a high-temperature gradient at greater depth²⁶, but there is no indication of a second, deeper horizontal velocity discontinuity in our migrations. Forward modelling with reflectivity synthetics establishes that a reflected phase from a horizontal discontinuity with a 5–10% shear velocity decrease is needed to produce reflections large enough to be observed in individual traces (see Supplementary Information). Thus any second phase-transformation feature is predicted²⁷ to be below the noise level of these and earlier¹³ migration images. Some strong negative-amplitude arrivals are apparent in the individual waveforms but these have travel times suggesting an out-of-plane scattering origin. Focusing by a three-dimensional low-velocity structure can produce comparable arrivals with a less pronounced velocity decrease. Lacking constraints on the scattering geometry, we cannot fully quantify the velocity anomaly involved, but an irregular structure with 3–10% low velocities and optimal focusing could account for the observed arrivals.

These three-dimensional low-velocity features at the edges of the shallower discontinuity image may be associated with lateral gradients in temperature on the margin of slab material and a resulting semi-vertical boundary to the phase change. Both features are imaged by scattered arrivals that arrive before and after ScS: such phases will not strongly illuminate a near-horizontal reflector, so this

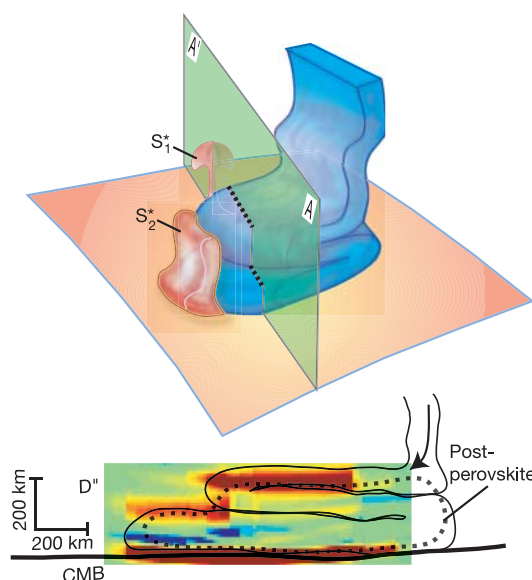


Figure 4 | A cold buckled subducted slab (blue) in the lowermost mantle may account for the thermal structure that results in a step in the perovskite/post-perovskite phase transition. Low-velocity scattering structures (S_1^* , S_2^*) on the lateral margins of the slab pile are postulated to account for the concentrated scatterers that are imaged. The green plane represents the profile A–A' through the migration volume used in Figs 1–3. The image in the lower panel is taken from profile A–A' in Fig. 2a, with positive-amplitude reflectivity (red) and negative-amplitude reflectivity (blue).

is likely to involve concentrated volumetric scatterers. Because any lateral velocity gradients from thermal and phase change effects are going to involve a reduction in velocity of only a few per cent, an additional contribution is needed to concentrate the heterogeneity. One viable candidate is small boundary-layer instabilities or plumes caused by slab material descending upon the CMB, extruding the hot basal layer of the thermal boundary layer, and prompting upwelling instabilities²⁵ of partial melt with strong velocity reductions. A lateral gradient of composition, perhaps of iron or aluminium content, could augment any thermal effects on the depth of the phase boundary⁶. However, for isolated chemical compositions to maintain such a steep slope, a very active process must be present to sustain it.

Our image volume is constrained by the path coverage, but the method could expand to cover a larger volume as more source/station geometries are included. The large step in the D'' reflector is comparable to features in continental lithosphere, which involve long-term thermal and chemical contrasts.

Received 23 December 2005; accepted 22 March 2006.

- Grand, S. P., van der Hilst, R. D. & Widiyantoro, S. Global seismic tomography: a snapshot of convection in the Earth. *GSA Today* **7**, 1–7 (1997).
- Grand, S. P. Mantle shear-wave tomography and the fate of subducted slabs. *Phil. Trans. R. Soc. Lond. A* **360**, 2475–2491 (2002).
- Lay, T. & Garnero, E. J. in *The State of the Planet: Frontiers and Challenges in Geophysics* (eds Sparks, R. S. J. & Hawkesworth, C. J.) 25–41 (Geophysical Monograph 150, American Geophysical Union, Washington DC, 2004).
- Wyssession, M. E., et al. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wyssession, M. E., Knittle, E. & Buffett, B. A.) 273–298 (Geodynamics Series 28, American Geophysical Union, Washington DC, 1998).
- Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. Post-perovskite phase transition in MgSiO₃. *Science* **304**, 855–858 (2004).
- Lay, T. et al. Multidisciplinary impact of the deep mantle phase transition in perovskite structure. *Eos* **86**, 1, 5 (2005).
- Masters, G., Laske, G., Bolton, H. & Dziewonski, A. M. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*, (eds Karato, S., Forte, A. M., Liebermann, R. C., Masters, G. & Stixrude, L.) 63–87 (American Geophysical Union, Washington DC, 2000).
- Wen, L., Silver, P., James, D. & Kuehnel, R. Seismic evidence for a thermo-chemical boundary at the base of the Earth's mantle. *Earth Planet. Sci. Lett.* **189**, 141–153 (2001).
- Ni, S. & Helmberger, D. V. Ridge-like lower mantle structure beneath South Africa. *J. Geophys. Res.* **108**, doi:10.1029/2001JB001545 (2003).
- Oganov, A. R. & Ono, S. Theoretical and experimental evidence for a post-perovskite phase of MgSiO₃ in Earth's D'' layer. *Nature* **430**, 445–448 (2004).
- Lay, T., Garnero, E. J. & Russell, S. A. Lateral variation of the D'' discontinuity beneath the Cocos Plate. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020300 (2004).
- Reasoner, C. & Revenaugh, J. Short-period P wave constraints on D'' reflectivity. *J. Geophys. Res.* **104**, 955–961 (1999).
- Thomas, C., Garnero, E. J. & Lay, T. High-resolution imaging of lowermost mantle structure under the Cocos Plate. *J. Geophys. Res.* **109**, doi:10.1029/2004JB003013 (2004).
- Ding, X. & Helmberger, D. V. Modeling D'' structure beneath Central America with broadband seismic data. *Phys. Earth Planet. Inter.* **101**, 245–270 (1997).
- Rokosky, J. M., Lay, T., Garnero, E. J. & Russell, S. A. High-resolution investigation of shear wave anisotropy in D'' beneath the Cocos Plate. *Geophys. Res. Lett.* **31**, L07605, doi:10.1029/2003GL018902 (2004).
- Lay, T. & Young, C. J. Imaging scattering structures in the lower mantle by migration of long-period S waves. *J. Geophys. Res.* **101**, 20023–20040 (1996).
- Revenaugh, J. A scattered-wave image of subduction beneath the Transverse Ranges, California. *Science* **268**, 1888–1892 (1995).
- Thomas, C., Weber, M., Wicks, C. W. & Scherbaum, F. Small scatterers in the lower mantle observed at German broadband arrays. *J. Geophys. Res.* **104**, 15073–15088 (1999).
- Kito, T., Krüger, F. & Negishi, H. Seismic heterogeneous structure in the lowermost mantle beneath the southwestern Pacific. *J. Geophys. Res.* **109**, B09304, doi:10.1029/2003JB002677 (2004).
- Lay, T. & Helmberger, D. V. A lower mantle S-wave triplication and the shear velocity structure of D. *Geophys. J. R. Astron. Soc.* **75**, 799–837 (1983).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
- Megnin, C. & Romanowicz, B. The three-dimensional shear velocity structure of the mantle from the joint inversion of body, surface, and higher mode waveforms. *Geophys. J. Int.* **143**, 709–728 (2000).
- Ritsema, J. & Van Heijst, H. J. Seismic imaging of structural heterogeneity in Earth's mantle: evidence for large-scale mantle flow. *Sci. Prog. (New Haven)* **83**, 243–259 (2000).
- Hung, S.-H., Garnero, E. J., Chiao, L.-Y., Kuo, B.-Y. & Lay, T. Finite-frequency tomography of D'' shear velocity heterogeneity beneath the Caribbean. *J. Geophys. Res.* **110**, B07305, doi:10.1029/2004JB003373 (2005).
- Tan, E., Gurnis, M. & Han, L. Slabs in the lower mantle and their modulation of plume formation. *Geochem. Geophys. Geosyst.* **3**, 1067, doi:10.1029/2001GC000238 (2002).
- Hernlund, J. W., Thomas, C. & Tackley, P. J. A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* **434**, 882–886 (2005).
- Flores, C. & Lay, T. The trouble with seeing double. *Geophys. Res. Lett.* **32**, L24305, doi:10.1029/2005GL024366 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M. Thorne provided three-dimensional finite difference synthetics for versions of our step discontinuity structure. We also thank Q. Williams for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R.H. (ahutko@pmc.ucsc.edu).

Embryological evidence for developmental lability during early angiosperm evolution

William E. Friedman¹

Recent advances in angiosperm phylogeny reconstruction^{1–3}, palaeobotany^{4,5} and comparative organismic biology^{6–8} have provided the impetus for a major re-evaluation of the earliest phases of the diversification of flowering plants. We now know that within the first fifteen million years of angiosperm history, three major lineages of flowering plants—monocotyledons, eumagnoliids and eudicotyledons—were established⁵, and that within this window of time, tremendous variation in vegetative and floral characteristics evolved. Here I report on a novel type of embryo sac (angiosperm female gametophyte or haploid egg-producing structure) in *Amborella trichopoda*, the sole member of the most ancient extant angiosperm lineage. This is the first new pattern of embryo sac structure to be discovered among angiosperms in well over half a century. This discovery also supports the emerging view^{9–12} that the earliest phases of angiosperm evolution were characterized by an extensive degree of developmental experimentation and structural lability, and may provide evidence of a critical link to the gymnospermous ancestors of flowering plants.

Nearly 130 years after Darwin proclaimed the origin and early evolution of flowering plants an “abominable mystery” (Charles Darwin’s letter to Joseph Dalton Hooker, 23 July 1879), reconstructing the biological features of the first angiosperms and linking them to non-flowering seed plant ancestors continue to challenge evolutionary biologists. Because flowering plants are distinguished from all other plants by diverse aspects of their reproductive biology, the importance of understanding the embryological features of early angiosperm lineages cannot be overstated. For more than a century, the presence of a highly reduced female gametophyte, within which a process of double fertilization initiates an embryo and a sexually formed embryo-nourishing tissue called endosperm, has been viewed as one of a suite of key characters integrally associated with the ‘success’ of angiosperms^{13,14}. Nevertheless, the origin and early evolutionary diversification of the angiosperm female gametophyte has remained enigmatic.

In the overwhelming majority of angiosperms, the female gametophyte contains seven cells and eight genetically identical nuclei at maturity and is referred to as the ‘Polygonum-type’ (Fig. 1). At the chalazal pole of a Polygonum-type female gametophyte are three sterile cells, the antipodals. At the micropylar pole, a three-celled egg apparatus contains the egg cell and two sterile accessory cells, the synergids. One synergid functions to accept the pollen tube and the initial deposition into the female gametophyte of the two sperm cells that will participate in the double fertilization process. The large central cell, which is the target of the unique angiosperm second fertilization event, is binucleate, and as a consequence, endosperm derived from a Polygonum-type female gametophyte is triploid. Given the near-ubiquity of reports of Polygonum-type female gametophytes among putatively ancient lineages, the conclusion was reached early and often in the twentieth century that the first

flowering plants must have had a seven-celled and eight-nucleate female gametophyte^{13–19} that initiated a triploid endosperm. In 1999, a series of molecular phylogenetic analyses revealed that many of the century-old assumptions about what constitute the most ancient angiosperm lineages were fundamentally incorrect^{20–23}. These and almost all subsequent phylogenetic analyses^{1–3} demonstrate that the monotypic New Caledonian taxon *Amborella trichopoda*, Nymphaeales (water lilies) and Austrobaileyales (Illiciaceae, Schisandraceae, Trimeniaceae and Austrobaileyaceae) comprise a basal-most grade of flowering plants whose origins predate the establishment of monocotyledons, eumagnoliids and eudicotyledons. More specifically, these analyses provide compelling evidence that *Amborella* (or *Amborella* plus Nymphaeales) is sister to all other angiosperms, and hence is one of the most ancient lineages of flowering plants. As such, *Amborella* may hold the key to deciphering the earliest phases of angiosperm evolutionary history and character transformation.

In order to examine the early evolution of gametic structures and the fertilization process, I performed a developmental analysis of the female gametophyte in *Amborella trichopoda*. Brightfield, fluorescence, confocal and transmission electron microscopy of *Amborella* ovules (field and greenhouse collections) clearly show that the mature female gametophyte comprises nine nuclei and eight cells. The egg apparatus contains a typical egg cell and three sterile uninucleate cells that share an interface with the egg cell (Figs 2 and 3; see also Supplementary Video). The large central cell is initially binucleate, and there are three antipodal cells that persist through the time of fertilization. Just before the fertilization process, the two polar nuclei of the central cell fuse into a single secondary nucleus (as is characteristic of many angiosperms) that is the target of the second fertilization event; triploid endosperm results (microspectrofluorometric DNA quantification data not shown). This basic structure of

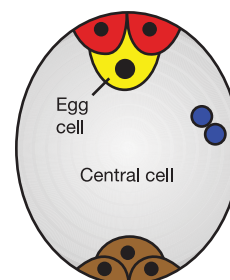


Figure 1 | A seven-celled, eight-nucleate female gametophyte (Polygonum-type) that is common to the vast majority of angiosperms. During the process of angiosperm double fertilization, the egg cell (yellow) and the central cell (grey) with two polar nuclei (blue) are each fertilized by a sperm cell and respectively produce a diploid zygote and a triploid embryo-nourishing tissue (endosperm). Synergid cells, red; antipodal cells, brown.

¹Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA.

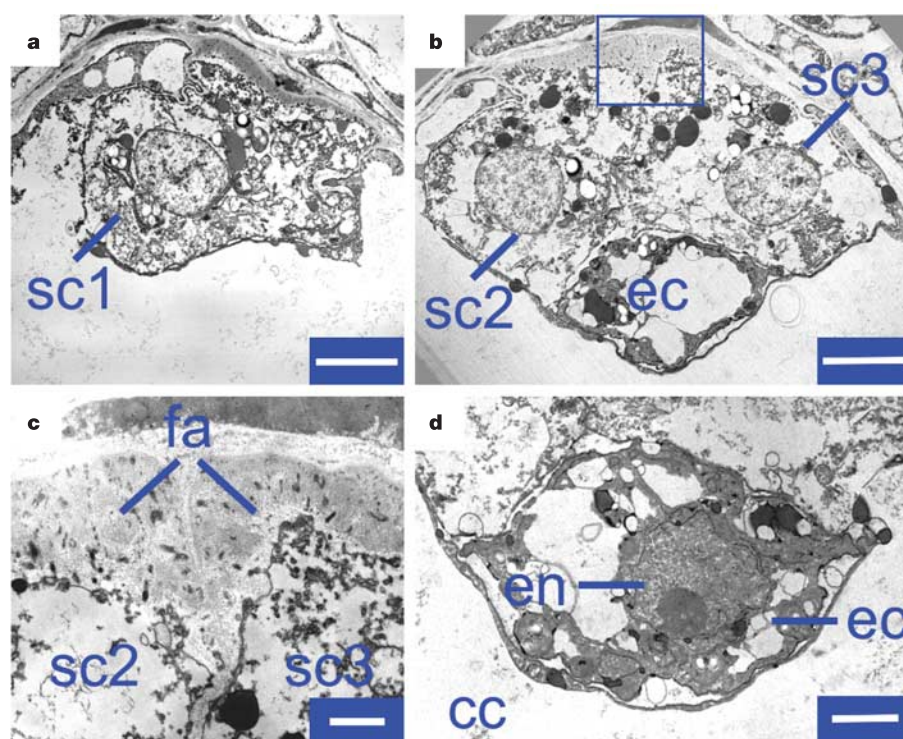


Figure 2 | Transmission electron micrographs of an individual serially sectioned female gametophyte of *Amborella*. **a–d**, The egg apparatus contains three synergid cells (sc1, sc2, sc3), each with a distinct nucleus (**a**, **b**), and an egg cell (ec), also with a distinct nucleus (en) (**d**). A filiform

apparatus (fa) is present in each of the three synergid cells of the egg apparatus and is shown (boxed area in **b**) at higher magnification in **c**. cc, central cell. Scale bars: **a**, **b**, 5 μ m; **c**, 1 μ m; **d**, 2.5 μ m.

the female gametophyte is characteristic of the more than 220 serially sectioned ovules (and seeds) that were examined.

Ultrastructural analysis of the mature female gametophyte demonstrates that each of the three sterile cells in the egg apparatus has a highly convoluted cell wall (similar to a transfer cell wall) at the micropylar pole that is characteristic of a standard filiform apparatus. Each sterile cell also shares a thin common wall with the pyramidal egg cell (Figs 2, 3). The filiform apparatus, as well as the cytoplasmic contents of the three sterile cells in the egg apparatus of *Amborella*, are essentially similar to each other and resemble those of synergids in other angiosperms that have been examined²⁴. Thus, the 'extra' cell in the egg apparatus is a synergid.

Female gametophyte development in *Amborella* parallels that of Polygonum-type female gametophytes until the terminal stage of differentiation of the egg apparatus (Figs 4, 5). Mitotic division of the one-nucleate female gametophyte yields two nuclei, one of which migrates to the chalazal pole. A second wave of mitosis results in two pairs of nuclei at opposite poles of the female gametophyte. A syncytial eight-nucleate stage was never observed, suggesting that this phase is extremely transitory (as is typically the case in other angiosperms²⁴); however, a seven-celled, eight-nucleate stage with three cells at each pole of the female gametophyte and a large binucleate central cell was seen in developing (pre-anthesis) floral buds (Fig. 4). At this stage, the female gametophyte of *Amborella* resembles the basic structure of a mature Polygonum-type female gametophyte: there are three antipodal cells at the chalazal pole, a binucleate central cell and three cells at the micropylar pole. In *Amborella*, however, the eight-nucleate, seven-celled female gametophyte represents the penultimate stage of the ontogeny. A single mitotic and cytokinetic event at the micropylar end of the *Amborella* female gametophyte produces the ninth nucleus and eighth cell (Fig. 4). The four cells at the micropylar pole, as well as the three antipodal cells at the chalazal pole, are initially flat, but increase in overall size as the female gametophyte matures (Figs 4, 5).

A four-celled egg apparatus is a definitive and characteristic feature of the female gametophyte of *Amborella* and cannot be viewed as a teratology. Among 147 serially sectioned mature *Amborella* female gametophytes (field and greenhouse collections) that were carefully examined, none could be shown to contain a three-celled egg apparatus.

The terminal cell division in the female gametophyte of *Amborella* produces the egg cell and the third synergid (Fig. 4). This can be inferred from examination of immature female gametophytes with three cells at the micropylar pole: each of the three cells has a face that abuts the boundary wall of the embryo sac (Fig. 4). In light of the fact that the egg cell only shares cell interfaces with the central cell and the three synergid cells (and not the wall of the embryo sac), the egg must be produced from a cell division that cleaves a daughter cell to the inside of the female gametophyte. Thus, the egg cell is the lineal sister cell of a synergid cell, and this developmental relationship is unique

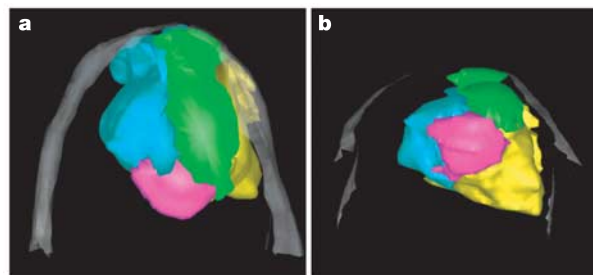


Figure 3 | Three-dimensional reconstructions (from semi-thin serial sections) of the four-celled egg apparatus of *Amborella*. **a**, All three sterile cells (synergids coloured green, blue and yellow) can be seen to abut the micropylar wall (at top) of the female gametophyte (grey) and envelope a portion of the egg cell (pink). **b**, The egg apparatus has been rotated 90° from **a** to show the pyramidal egg cell seated on all three synergids.

among all angiosperms. In all other angiosperm female gametophytes, the egg nucleus is the mitotic sister nucleus of a polar nucleus of the central cell²⁴.

The eight-celled, nine-nucleate female gametophyte of *Amborella* herein described is the first new type of embryo sac to be discovered among flowering plants in more than half a century of extensive surveys, and is the only one known to produce normally more than two synergids in the egg apparatus. *Indeterminate gametophyte1* mutants in maize frequently produce extra synergids within the egg apparatus, but this is always accompanied by supernumerary egg cells²⁵.

Until quite recently, *Amborella* and members of the Nymphaeales and Austrobaileyales were generally thought to exhibit a standard seven-celled, eight-nucleate Polygonum-type female gametophyte²⁶. However, earlier reports of Polygonum-type female gametophytes in Nymphaeales and Austrobaileyales have now been shown to be erroneous^{8,27–29}, as is also clearly the case for *Amborella*. Ironically, it is now evident that none of the most ancient lineages of flowering plants produces a seven-celled, eight-nucleate female gametophyte, a stark reminder that much remains to be discovered or correctly circumscribed for the earliest angiosperms. The mature female gametophytes of members of the Nymphaeales and Austrobaileyales contain four cells and four nuclei at maturity: a haploid uninucleate central cell, an egg cell and two synergids^{8,27}. The mature female gametophyte of *Amborella* contains eight cells and nine nuclei at maturity.

The issue of whether the eight-celled, nine-nucleate female gametophyte of *Amborella* represents an autapomorphy of the *Amborella* clade or is a plesiomorphic characteristic for all angiosperms cannot be resolved at this time. However, two (among several) evolutionary developmental explanations are particularly intriguing. One possibility is that the *Amborella*-type female gametophyte is peramorphic and derived from a late developmental modification of a standard Polygonum-type female gametophyte. As such, this would suggest that the Polygonum-type might be plesiomorphic for angiosperms, even though a seven-celled, eight-nucleate female gametophyte is not present in any of the most ancient extant lineages of flowering plants.

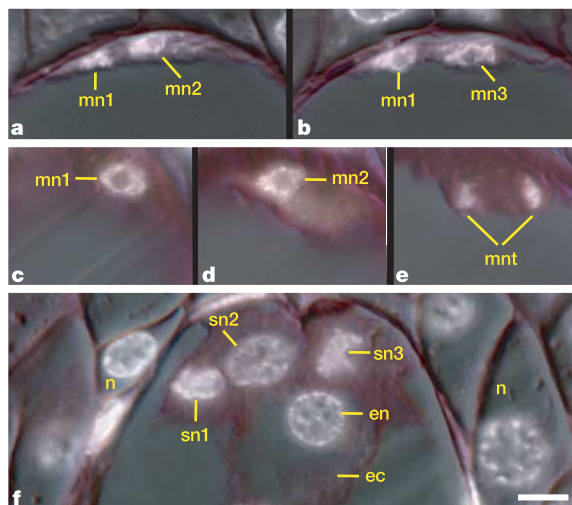


Figure 4 | Terminal developmental stages of the egg apparatus in the female gametophyte of *Amborella*. **a, b**, Combined fluorescent and differential interference contrast images of consecutive serial sections showing three DAPI-stained micropylar nuclei (mn1, mn2, mn3). A prominent nucleolus is characteristic of each female gametophyte nucleus at this stage. **c–e**, Fluorescent images of consecutive serial sections showing two DAPI-stained micropylar nuclei (mn1, mn2) and a telophase mitotic figure (mnt) that will yield the egg and a synergid. **f**, Mature egg apparatus with three synergids and nuclei (sn1, sn2, sn3) and the egg cell (ec) with egg nucleus (en). n, nucellus. Scale bar, 5 μ m.

According to this hypothesis, a terminal addition (a single cell division to create the egg and an additional synergid) to the ontogenetic sequence of the Polygonum-type would have led to the creation of the novel *Amborella*-type female gametophyte. Within the precepts of this hypothesis, the four-nucleate and four-celled female gametophytes of Nymphaeales and Austrobaileyales would be viewed as derived (and not plesiomorphic) within angiosperms, a conclusion at odds with that recently articulated by Williams and Friedman²⁷ and Friedman and Williams^{8,29}.

Alternatively, the unique four-celled egg apparatus in *Amborella* could represent a critical link between angiosperms and gymnosperms. An important difference between female gametogenesis in gymnosperms and angiosperms relates to the fact that gymnosperms never directly form an egg cell; rather they form a mother cell ('central cell' sensu gymnosperms) within an archegonium that later divides to yield a ventral canal cell (or nucleus, if no cytokinesis occurs) and an egg cell³⁰. In all flowering plants²⁴, with the exception of *Amborella*, egg cells are always directly produced at the time of cellularization of the syncytial female gametophyte. One way to account for the four-celled egg apparatus in *Amborella* is that the mother cell that divides to give rise to the egg cell and third synergid is the homologue of the egg mother cell (central cell sensu gymnosperms) within the archegonium of a non-flowering seed plant. In this interpretation, the third synergid cell can be viewed as homologous to the ventral canal cell of a gymnosperm archegonium. Ultimately, if it can be shown that *Amborella* contains vestiges of the basic archegonium of non-flowering plants, this discovery will certainly rise to the standard of a 'missing link'.

It must be noted that should the phylogenetic placement of *Amborella* as a member of the most ancient lineage of angiosperms shift in future systematic analyses, interpretations of the evolutionary developmental significance of the *Amborella*-type female gametophyte will have to be re-evaluated. Nevertheless, there are now three distinct fully documented ontogenetic sequences present among clades whose origins date to the earliest phases of angiosperm history: the eight-celled, nine-nucleate sequence of *Amborella*, the four-celled, four-nucleate sequence characteristic of Nymphaeales and Austrobaileyales^{27–29}, and the typical Polygonum-type sequence characteristic of over 70% of extant angiosperms²⁴, including basal monocotyledons, eumagnoliids and eudicotyledons²⁹. Given that only 5 yr ago *Amborella*, Nymphaeales and Austrobaileyales were widely (although not universally) believed to produce a standard Polygonum-type female gametophyte^{26,29}, these findings are a strong reminder that embryological features are far more diverse among ancient flowering plant lineages than had ever been anticipated. Finally, these data are also strongly congruent with other data sets^{4–12}

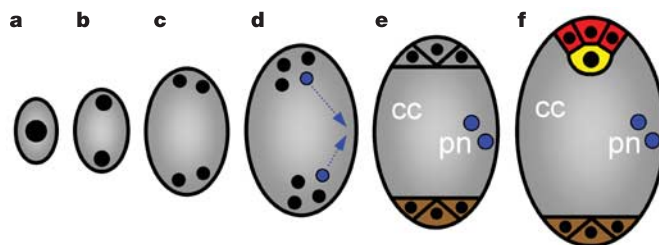


Figure 5 | Development of the female gametophyte of *Amborella*. **a–d**, A series of three free-nuclear mitoses yields an eight-nucleate syncytium (**d**). **e**, A brief phase of cellularization creates a structure with three antipodal cells (brown) at the chalazal pole, three cells at the micropylar pole and a binucleate central cell (cc) with two polar nuclei (pn) that resembles a mature *Polygonum*-type female gametophyte. **f**, A single cell division in one of the micropylar cells produces an egg cell (yellow) and a cell that differentiates into one of the three synergids (red) characteristic of the mature eight-celled, nine-nucleate female gametophyte. The micropylar pole is at the top.

that show that the earliest phases of flowering plant evolution were marked by a tremendous degree of structural lability and developmental experimentation.

METHODS

Plant collection. Reproductive material of *Amborella trichopoda* was collected from plants growing in the field in New Caledonia at Plateau de Dogny (GPS coordinates = 21° 37.253' S, 165° 52.130' E) and greenhouse populations at the University of Colorado. Supplemental pollinations were performed on a subset of flowers in the field and in greenhouses.

Brightfield and fluorescence microscopy. Carpellate flowers were chemically fixed for 24–48 h in either 3:1 95% ethanol:acetic acid or 4% glutaraldehyde in a solution of 50 mM PIPES buffer (also 5 mM EGTA and 1 mM MgSO₄, pH 6.8) or 0.02 M potassium phosphate buffer. Specimens were dehydrated through an ethanol series, infiltrated and embedded with glycol methacrylate, and serially sectioned into 5- or 3-µm-thick ribbons. Flowers fixed in 3:1 95% ethanol:acetic acid were stained with a solution of 0.25 µg ml⁻¹ of 4',6-diamidino-2-phenylindole (DAPI) and 0.1 mg ml⁻¹ *p*-phenylenediamine in 0.05 M Tris (pH 7.2). Flowers fixed in glutaraldehyde were stained in 0.1% toluidine blue or periodic acid Schiff's reaction and DAPI. Fluorescence and brightfield digital micrographs were taken on a Zeiss Axiophot microscope equipped with a Zeiss AxioCam digital camera, and images were processed with Adobe Photoshop software. For visualization of DAPI, ultraviolet filter set (model no. 48702) with excitation filter (365 nm, band-pass 12 nm), dichroic mirror (FT395), and barrier filter (LP397) was used with a Zeiss Plan Neofluar 63× objective.

Transmission electron microscopy. Carpellate flowers were chemically fixed for 24–48 h in 4% glutaraldehyde in 0.02 M potassium phosphate buffer or a solution of 2% paraformaldehyde, 1% glutaraldehyde and 2% acrolein in 50 mM PIPES buffer (pH 6.8). Ovules were dissected out of individual carpels and post-fixed with 2% osmium tetroxide at 4 °C for 2 h, rinsed, dehydrated through a standard ethanol series and infiltrated with propylene oxide for 1.5 h. Tissues were then infiltrated with Spurr's resin and embedded. Female gametophytes were serially sectioned with a diamond knife at a thickness of 70 nm, mounted on 200 mesh copper grids, stained with lead citrate, and examined on a Phillips CM 10 transmission electron microscope.

Confocal microscopy. Prepared slides stained with DAPI and periodic acid Schiff's reagent were examined with a Leica TCS SP2 laser scanning confocal microscope and excited with a 405 nm blue diode. DAPI emission was detected within the 440–500 nm range, Schiff's reagent in the 570–630 nm range and autofluorescence in the 650–690 nm range. All images were captured with a Leica oil-immersion 63× objective.

Three-dimensional reconstructions. Tissues embedded in Spurr's resin were serially sectioned with a glass knife at a thickness of 0.5 µm, stained with toluidine blue and digitally photographed (see brightfield microscopy). Outlines of each cell of the egg apparatus, as well as the micropylar portion of the female gametophyte wall, were digitized and serially aligned in IMOD (Copyright, Boulder Laboratory for 3-D Electron Microscopy of Cells). IMOD programs were then used to reconstruct and display three-dimensional images of the egg apparatus.

Received 9 January; accepted 1 March 2006.

1. Soltis, P. S. & Soltis, D. E. The origin and diversification of angiosperms. *Am. J. Bot.* **91**, 1614–1626 (2004).
2. Leebens-Mack, J. *et al.* Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol. Biol. Evol.* **22**, 1948–1963 (2005).
3. Qiu, Y.-L. *et al.* Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *Int. J. Plant Sci.* **166**, 815–842 (2005).
4. Friis, E. M., Pedersen, K. R. & Crane, P. R. Fossil floral structures of a basal angiosperm with monocolpate, reticulate-acuminate pollen from the Early Cretaceous of Portugal. *Grana* **39**, 226–239 (2000).
5. Crane, P. R., Herendeen, P. & Friis, E. M. Fossils and plant phylogeny. *Am. J. Bot.* **91**, 1683–1699 (2004).
6. Carlquist, S. & Schneider, E. L. The tracheid-vessel element transition in angiosperms involves multiple independent features: Cladistic consequences. *Am. J. Bot.* **89**, 185–195 (2002).

7. Feild, T. S., Arens, N. C. & Dawson, T. E. The ancestral ecology of angiosperms: Emerging perspectives from extant basal lineages. *Int. J. Plant Sci.* **164**, S129–S142 (2003).
8. Friedman, W. E. & Williams, J. H. Developmental evolution of the sexual process in ancient flowering plant lineages. *Plant Cell* **16**, S119–S132 (2004).
9. Doyle, J. A. & Endress, P. K. Morphological phylogenetic analysis of basal angiosperms: Comparison and combination with molecular data. *Int. J. Plant Sci.* **161**, S121–S153 (2000).
10. Endress, P. K. Origins of flower morphology. *J. Exp. Zool.* **291**, 105–115 (2001).
11. De Craene, L. P. R., Soltis, P. S. & Soltis, D. E. Evolution of floral structures in basal angiosperms. *Int. J. Plant Sci.* **164**, S329–S363 (2003).
12. Zanis, M. J., Soltis, P. S., Qiu, Y. L., Zimmer, E. & Soltis, D. E. Phylogenetic analyses and perianth evolution in basal angiosperms. *Ann. Mo. Bot. Gard.* **90**, 129–150 (2003).
13. Stebbins, G. L. *Flowering Plants: Evolution Above the Species Level* (Harvard Univ. Press, Cambridge, 1974).
14. Cronquist, A. *The Evolution and Classification of Flowering Plants* 2nd edn (New York Bot. Gard., Bronx, 1988).
15. Porsch, O. *Versuch einer phylogenetischen Erklärung des Embryosackes und der doppelten Befruchtung der Angiospermen* (Gustav Fischer, Jena, 1907).
16. Schnarf, K. *Vergleichende Embryologie der Angiospermen* (Borntraeger, Berlin, 1931).
17. Johri, B. M. in *Recent Advances in the Embryology of Angiosperms* (ed. Maheshwari, P.) 69–103 (Int. Soc. Plant Morphol., Delhi, 1963).
18. Davis, G. L. *Systematic Embryology of the Angiosperms* (John Wiley and Sons, New York, 1966).
19. Donoghue, M. J. & Scheiner, S. M. in *Ecology and Evolution of Plant Reproduction* (ed. Wyatt, R.) 356–389 (Chapman & Hall, New York, 1992).
20. Mathews, S. & Donoghue, M. J. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947–950 (1999).
21. Parkinson, C. L., Adams, K. L. & Palmer, J. D. Multigene analyses identify the three earliest lineages of extant flowering plants. *Curr. Biol.* **9**, 1485–1488 (1999).
22. Qiu, Y.-L. *et al.* The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
23. Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a research tool for comparative biology. *Nature* **402**, 402–404 (1999).
24. Huang, B.-Q. & Russell, S. D. Female germ unit: Organization, isolation, and function. *Int. Rev. Cytol.* **140**, 233–293 (1992).
25. Guo, F. L., Huang, B. Q., Han, Y. Z. & Zee, S. Y. Fertilization in maize indeterminate gametophyte1 mutant. *Protoplasma* **223**, 111–120 (2004).
26. Tobe, H., Jaffre, T. & Raven, P. H. Embryology of *Amborella* (Amborellaceae): descriptions and polarity of character states. *J. Plant Res.* **113**, 271–280 (2000).
27. Williams, J. H. & Friedman, W. E. Identification of diploid endosperm in an early angiosperm lineage. *Nature* **415**, 522–526 (2002).
28. Friedman, W. E., Gallup, W. N. & Williams, J. H. Female gametophyte development in *Kadsura*: implications for Schisandraceae, Illiciaceae, and the early evolution of flowering plants. *Int. J. Plant Sci.* **164**, S293–S305 (2003).
29. Williams, J. H. & Friedman, W. E. The four-celled female gametophyte of *Illicium* (Illiciaceae; Austrobaileyales): implications for understanding the origin and early evolution of monocots, eumagnoliids, and eudicots. *Am. J. Bot.* **91**, 332–351 (2004).
30. Gifford, E. M. & Foster, A. S. *Morphology and Evolution of Vascular Plants* (W.H. Freeman and Co., New York, 1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank P. K. Diggle, R. H. Robichaux and L. Hufford for critical comments on the manuscript; K. C. Ryerson for assistance with all aspects of data collection and field work; and A. J. Redford, D. O'Connor, T. J. Lemieux and E. N. Madrid for assistance with various aspects of histology and field work. Field collections of *Amborella* in New Caledonia were made possible by permission of the Direction des Ressources Naturelles, Province Sud, Nouvelle-Calédonie and facilitated by B. Fogliani. This work was supported by grants from the National Science Foundation. To A.S.F., who read every paper I wrote.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to W.E.F. (ned@colorado.edu).

Copper-containing plastocyanin used for electron transport by an oceanic diatom

Graham Peers¹ & Neil M. Price¹

The supply of some essential metals to pelagic ecosystems is less than the demand, so many phytoplankton have slow rates of photosynthetic production and restricted growth¹. The types and amounts of metals required by phytoplankton depends on their evolutionary history² and on their adaptations to metal availability^{3,4}, which varies widely among ocean habitats. Diatoms, for example, need considerably less iron (Fe) to grow than chlorophyll-*b*-containing taxa², and the oceanic species demand roughly one-tenth the amount of coastal strains^{5–7}. Like Fe, copper (Cu) is scarce in the open sea, but notably higher concentrations of it are required for the growth of oceanic than of coastal isolates⁸. Here we report that the greater Cu requirement in an oceanic diatom, *Thalassiosira oceanica*, is entirely due to a single Cu-containing protein, plastocyanin, which—until now—was only known to exist in organisms with chlorophyll *b* and cyanobacteria. Algae containing chlorophyll *c*, including the closely related coastal species *T. weissflogii*, are thought to lack plastocyanin and contain a functionally equivalent Fe-containing homologue, cytochrome *c*₆ (ref. 9). Copper deficiency in *T. oceanica* inhibits electron transport regardless of Fe status, implying a constitutive role for plastocyanin in the light reactions of photosynthesis in this species. The results suggest that selection pressure imposed by Fe limitation has resulted in the use of a Cu protein for photosynthesis in an oceanic diatom. This biochemical switch reduces the need for Fe and increases the requirement for Cu, which is relatively more abundant in the open sea.

Offshore oceanic waters contain extremely low concentrations of dissolved Fe, manganese (Mn) and zinc (Zn), and phytoplankton in these regions have adapted by reducing the amounts of these metals needed for growth. Coastal species thrive in comparatively metal-enriched habitats, and have a much greater need for these resources because their metabolic requirements are higher¹⁰. Like other biologically active metals, Cu shows an offshore–onshore gradient, with total dissolved concentrations ranging between 0.4 nM in oceanic surface waters¹¹ and 50 nM in unpolluted coastal bays and estuaries¹². Although Cu concentration is lower in the open sea, oceanic phytoplankton need much higher concentrations of it to grow than coastal species⁸. This observation is opposite to that reported previously for other biologically active metals^{6,10,13} (Supplementary Fig. 1).

The greater need for dissolved Cu in one oceanic strain is due to a higher metabolic requirement for growth (Fig. 1). Under mildly Cu-limiting conditions, the intracellular Cu quota of *T. oceanica* (isolated from the metal-impooverished Sargasso Sea) is ten times greater than that of a comparably Cu-deficient coastal strain, *T. weissflogii* (2.76 and 3.53 $\mu\text{mol Cu}$ per litre cell volume compared to $0.28 \pm 0.16 \mu\text{mol Cu}$ per litre cell volume; $n = 2$ and $n = 4$, respectively; values are mean $\pm$ s.d.). These values represent the first true estimates of metabolic Cu in diatoms and are roughly 10–20% of the cellular values reported previously¹⁴.

Most of the Fe, Mn and Cu required by photoautotrophs is found in electron transport proteins and enzymes associated with photosynthesis⁹. Iron-containing cytochrome *c*₆ is believed to be the sole electron transport protein between cytochrome *b*₆/*f* and photosystem I (PSI) in diatoms and other chlorophyll-*c*-containing algae⁹. Previous studies have found it in *T. weissflogii*^{4,15}, and we confirmed its presence in this species by its reduced-minus-oxidized spectral peak (the change in absorbance between reduced and oxidized states at different wavelengths) at $\sim 554 \text{ nm}$ (Supplementary Fig. 2). Cytochrome *c*₆ was not present in similarly prepared soluble protein extracts from *T. oceanica* (Supplementary Fig. 2 and ref. 4), which instead contained a chromophore with an oxidized-minus-reduced spectrum corresponding to a type I Cu protein (Supplementary Fig. 2 and Fig. 2b). The type I Cu protein was purified to homogeneity and a portion of it was sequenced. It has an apparent molecular weight of $\sim 18.2 \text{ kDa}$ (Fig. 2a) and an amino-terminal amino-acid composition that is very similar to plastocyanins from cyanobacteria (Fig. 2c). An internal sequence of the protein contains one of the conserved Cu-binding domains, which most closely resembles the plastocyanins of higher plants and green algae (sharing 82% identity with plastocyanin from *Arabidopsis thaliana*; Supplementary Fig. 3). Thus, our results show that *T. oceanica* synthesizes a Cu-containing plastocyanin.

In higher plants, green algae and cyanobacteria, plastocyanin transfers electrons from the cytochrome *b*₆/*f* complex to PSI. If it performs the same function in *T. oceanica*, then we expect that photosynthetic electron transport would be inhibited when cells

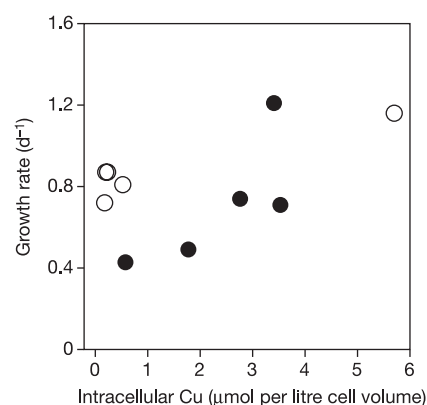


Figure 1 | Growth rate of *Thalassiosira* as a function of intracellular Cu concentration. *T. oceanica* (filled circles) and *T. weissflogii* (open circles) were grown in either Cu-sufficient media (containing 21.4 nM Cu) or in Cu-deficient media (containing no added Cu for *T. weissflogii*, and either 1.0 or 2.5 nM Cu for *T. oceanica*), and the intracellular Cu concentrations measured.

¹Department of Biology, McGill University, 1205 Avenue Dr Penfield, Montreal, Quebec H3A 1B1, Canada.

are Cu-deficient. We used photosystem II (PSII) fluorescence emission techniques to probe the various redox transitions associated with primary photochemistry in PSII and the photosynthetic electron carriers¹⁶. The maximum potential efficiency of PSII (F_v/F_m) assessed by pulse-amplitude modulated (PAM) fluorescence was identical in Cu-limited and control cells (0.55/0.55 and 0.55/0.56, respectively; $n = 2$; see also Supplementary Table 1). Therefore, PSII integrity and function *per se* are independent of Cu nutrition. However, the realized efficiency of PSII, which measures the operational efficiency under actinic light (approximating growth irradiance), was reduced under low Cu conditions (0.12/0.14 compared with 0.31/0.33 for metal-replete conditions; $n = 2$). Thus, electron flow downstream of PSII is impeded by a lack of Cu. The fluorescence transient in *T. oceanica* is strongly affected by Cu-deficiency (Fig. 3a) and is consistent with this interpretation. Copper limitation decreased the time to reach peak fluorescence (F_p) in dark-adapted cells when the plastoquinone pool was fully reduced and increased the duration of F_p from 300 ± 10 to $2,200 \pm 100$ ms ($n = 3$). These results suggest that oxidation of the plastoquinone pool is impeded. PSII fluorescence decay kinetics also verify that the rate of electron transfer from PSII to PSI is decreased in Cu-limited cells (Supplementary Table 1). Similar PSII fluorescence patterns are seen in mutants of *A. thaliana* lacking active plastocyanin¹⁷. Our fluorescence data are consistent with plastocyanin having an active role in photosynthetic electron transport in *T. oceanica*.

When Cu is low, *T. oceanica* seems to be unable to assemble adequate amounts of functional, Cu-containing plastocyanin, thus limiting electron flow. Although many plastocyanin-containing algae are known to synthesize cytochrome c_6 under Cu deficiency¹⁸, our data suggest that this is unlikely in *T. oceanica*, because photosynthetic electron flow is unaltered by Fe in the presence of low Cu (data not shown). Copper limitation has little effect on the fluorescence transient of the coastal species, as expected owing to its apparent lack of plastocyanin (Fig. 3b). This result is consistent with earlier

reports that were unable to detect this protein in *T. weissflogii*^{4,15}. We note that a closely related species, *T. pseudonana*, lacks the gene for plastocyanin (see <http://genome.jgi-psf.org/thaps1/thaps1.home.html>) and has similar Cu requirements to *T. weissflogii*⁸. Thus, we propose that *T. oceanica* and *T. weissflogii* use different redox carriers for the transport of electrons between cytochrome b_6/f and PSI.

Using a molar absorption coefficient of $4.5 \text{ mM}^{-1} \text{ cm}^{-1}$ for a type I Cu protein¹⁸ and measurements of the total biomass used for partial purification, we calculate that *T. oceanica* contains $\sim 3 \mu\text{mol Cu}$ per litre cell volume in plastocyanin. This quantity of Cu is equivalent to the additional amount required by the oceanic species compared with the coastal strain ($2.9 \mu\text{mol Cu}$ per litre cell volume). Thus, the difference in Cu content between these two diatoms can be solely attributed to the presence of plastocyanin. On the basis of concentrations of PSI in *T. oceanica*, we estimate that the plastocyanin:PSI ratio is approximately 1.5:1, in agreement with the value reported in higher plants¹⁹ and the ratio of cytochrome c_6 :PSI in *T. weissflogii*⁴. The cellular Cu in *T. weissflogii* represents basal metabolic requirements for Cu within such enzymes as cytochrome c oxidase²⁰ and the multicopper oxidase used in Fe transport^{8,21}. These requirements are shared by *T. oceanica*, but are minor in comparison with its need for Cu in plastocyanin.

The two diatoms investigated here are closely related²²; hence, the use of plastocyanin by one species and of cytochrome c_6 by the other to perform the same function in the photosynthetic electron transport chain is quite surprising. In organisms containing both cytochrome c_6 and plastocyanin, transfer of electrons to PSI occurs at the same rate regardless of which carrier is functional¹⁹, so there should be no intrinsic advantage of using one protein over the other. A combination of gene loss events, possibly accompanied by lateral gene transfer of plastocyanin into the oceanic diatom, could account for the taxonomic distribution of these redox proteins²³. At this time, the evolutionary roots of this pattern are uncertain.

Selection for the use of plastocyanin or cytochrome c_6 in diatoms could be related to the availabilities of Fe and Cu in the environment. In the open sea, dissolved Fe concentration is extremely low (0.07 nM) and limiting to phytoplankton, whereas Cu is present in

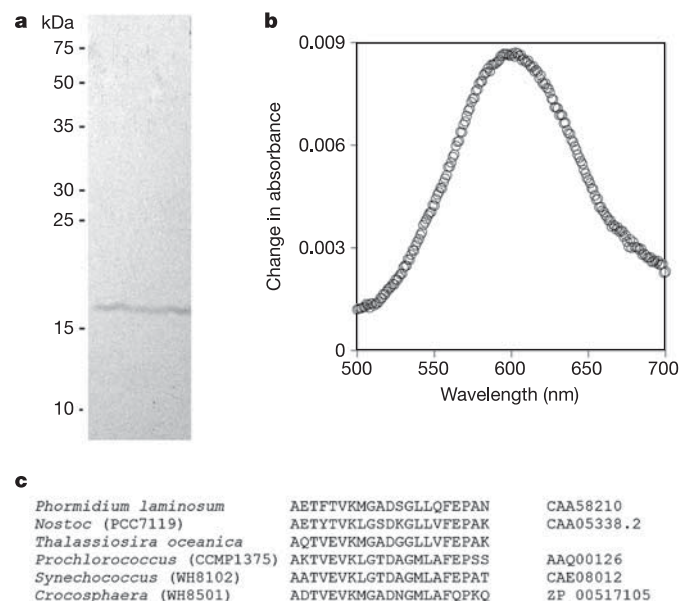


Figure 2 | Isolation of plastocyanin from the oceanic diatom *Thalassiosira oceanica*. **a**, Coomassie-blue staining of the purified protein resolved by denaturing polyacrylamide gel electrophoresis (SDS-PAGE; 15% polyacrylamide). The position and size of the molecular weight markers are shown to the left of the gel. **b**, Oxidized-minus-reduced spectrum of the purified protein showing a maximum absorbance change peak at $\sim 600 \text{ nm}$. **c**, The N-terminal amino-acid sequence of the purified protein aligned (using the ClustalW program; <http://www.ebi.ac.uk/clustalw/>) with plastocyanin sequences of cyanobacteria. Accession numbers for the cyanobacterial sequences are shown.

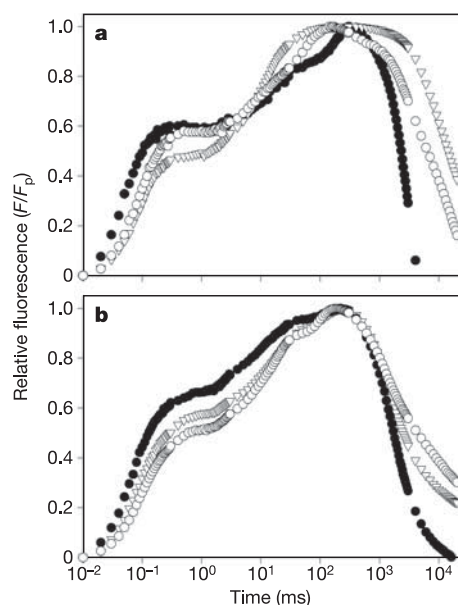


Figure 3 | PSII fluorescence transients of *Thalassiosira*. **a**, *T. oceanica* and **b**, *T. weissflogii* were grown in Cu-deficient (open triangles), Fe-deficient (filled circles) or metal-replete (open circles) media. Initial fluorescence values were set at zero, and fluorescence values were normalized to peak fluorescence (F_p) for ease of comparison.

higher concentrations (0.4–1 nM)¹¹. Functional assembly of these redox proteins depends on an adequate supply of the metal cofactors, so under these conditions plastocyanin synthesis should be favoured. Conversely, diatoms in Fe-replete coastal waters should have sufficient Fe for synthesis of cytochrome *c*₆. The use of plastocyanin by oceanic species would have the added benefit of reducing the amount of Fe required for growth. Switching from cytochrome *c*₆ to plastocyanin reduces cellular Fe demand by roughly 10% based on the plastocyanin concentrations reported above and an Fe quota of 30 $\mu\text{mol Fe}$ per litre cell volume (ref. 7). Any Fe saved by these substitutions could be allocated to other essential functions, so the advantages of using plastocyanin over cytochrome *c*₆ should be significant. *T. oceanica* is also known to have greatly reduced its Fe requirements for growth by decreasing the cellular concentrations of cytochrome *b*₆/*f* and PSI⁴, and by synthesizing flavodoxin instead of ferredoxin²⁴. Despite such changes to the photosynthetic apparatus, *T. oceanica* maintains rapid electron transport rates under Fe-depleted conditions²⁴. A substantial ratio of soluble electron carriers to PSI is thus needed to allow rapid turnover of the electron transport chain¹⁹. The use of plastocyanin reported in this study may allow *T. oceanica* to maintain fast electron transfer rates despite the high ratios of PSII:PSI observed in this organism during Fe-replete and Fe-depleted growth²⁴.

Synthesis of plastocyanin by the oceanic diatom increases the cellular demand for Cu by tenfold compared with the coastal species. Because Cu uptake is strongly dependent on Cu concentration, and organic ligands^{25,26} reduce the reactivity of Cu and alter its availability to phytoplankton, is the level of biologically available Cu in the open ocean high enough to meet the requirements for diatom growth? It would seem so, judging by the much greater Cu:C ratios of phytoplankton in the field (1.7–3.6 and 4.0 $\mu\text{mol mol}^{-1}$; refs 27 and 14, respectively) than in our Cu-limited cultures (0.24 $\mu\text{mol mol}^{-1}$). These field estimates, from the North Pacific Ocean, show that Cu levels are sufficient to allow diatoms to acquire ten times the amount of Cu necessary for plastocyanin synthesis. Although Cu levels may at times be limiting⁸, dissolved measurements from other oceans²⁵ show similar amounts of Cu as in the North Pacific, suggesting that Cu is generally plentiful. Thus, the use of plastocyanin instead of cytochrome *c*₆ in photosynthesis is a viable strategy that minimizes diatom dependence on Fe and helps maximize fitness in an Fe-deficient ocean.

METHODS

Cultures. Axenic cultures of *T. oceanica* (strain CCMP 1005) and *T. weissflogii* (strain CCMP 1336) were obtained from the CCMP (Center for Culture of Marine Phytoplankton). Cultures were grown in artificial seawater medium (AQUIL) in acid-cleaned polycarbonate flasks at 20 °C and under 180 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ light supplied by cool white fluorescent bulbs⁸. Growth rates, cell volumes and cell densities were determined as previously described²⁸.

Copper quotas. Diatom cultures were grown in semicontinuous batch cultures in 0.5 litres of medium containing 21.4, 2.5, 1.0 or 0 nM Cu and a full complement of all other trace metals. Diatoms were fully acclimated (<10% variation in growth between successive transfers in each treatment) before analysis. Cells were collected onto acid-cleaned (10% TraceMetal-grade HCl, Fisher Scientific) 25-mm diameter, 2- μm pore size polycarbonate filters housed in an acid-cleaned polysulphone filtration apparatus using a gentle vacuum. Collected cells were rinsed with 25 ml of AQUIL medium containing 100 μM EDTA with no added metals. Diatoms were digested as described²⁹ and Cu analysed by atomic absorption spectroscopy (Perkin Elmer AAnalyst 800 spectrometer equipped with a Zeeman graphite furnace). Filter blanks, treated as above, were below detection limits. Copper concentrations in the digest were normalized to total filtered cell volume.

Chlorophyll fluorescence. The fluorescence transient of PSII was measured using a Hansatech Plant Efficiency Analyser (PEA). Mid-exponential growth phase cultures were dark acclimated for at least 15 min before filtering 30–60 ml of culture onto Millipore AP20 glass fibre filters. The filter was immediately placed into a leaf-disk adaptor supplied by the manufacturer and exposed to a saturating pulse of red light (650 nm) at an intensity of 3,000 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ for 20 s. Fluorescence readings are reported as the fluorescence (*F*) normalized to the

peak signal (*F*_p). Other fluorescence induction parameters were measured using a pulse-amplitude modulated fluorometer (Water PAM, Heinz Walz). Diatoms were concentrated from mid-exponential cultures by filtration and immediately resuspended in approximately 5 ml of AQUIL medium and dark acclimated for at least 10 min. The maximum quantum efficiency of PSII (*F*_v/*F*_m) was measured using a saturating pulse of 3,000 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ supplied for 600 ms. Actinic illumination of samples with 150 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ of light was used for the measurement of the realized efficiency of PSII (Φ_{PSII} ; both parameters were calculated according to published methods⁴).

Protein purification. Diatoms were grown under metal-replete conditions in 20-litre carboys with vigorous aeration. Late-exponential-phase cultures were collected by positive pressure onto 142-mm diameter, 3- μm pore size polycarbonate filters. Cells were quickly removed from the filters and stored in liquid nitrogen until further handling. Frozen cells were resuspended in 50 mM phosphate buffer, pH 7.0, and the slurry passed through an ice-cold Yeda press (three times). NaCl (50 mM, final concentration) was added and the slurry then sonicated for 1 min on ice using a Branson sonicator equipped with a micro-tip probe. Protein concentrations in the initial homogenates were 214 mg ml^{-1} and 2.25 mg ml^{-1} for *T. oceanica* and *T. weissflogii*, respectively (BCA assay, Sigma). Following selective precipitations³⁰, the resulting protein preparation was then centrifuged for 10 min at 16,000g and the soluble protein was analysed spectrophotometrically (Cary 1E spectrophotometer, 1-cm pathlength quartz cuvette). We continued with the purification protocol to isolate plastocyanin from *T. oceanica*. The purified protein was subjected to two-dimensional (2D) gel electrophoresis followed by N-terminal sequencing via Edman degradation using standard protocols. Additional sequences were obtained from trypsin-digested protein.

Received 7 January; accepted 7 February 2006.

Published online 29 March 2006.

- Morel, F. M. M. & Price, N. M. The biogeochemical cycles of trace metals in the oceans. *Science* **300**, 944–947 (2003).
- Quigg, A. *et al.* The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature* **425**, 291–293 (2003).
- Cullen, J. T., Lane, T. W., Morel, F. M. M. & Sherrell, R. M. Modulation of cadmium uptake in phytoplankton by seawater CO₂ concentration. *Nature* **402**, 165–167 (1999).
- Strzepek, R. F. & Harrison, P. J. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* **431**, 689–692 (2004).
- Ryther, J. H. & Kramer, D. D. Relative iron requirements of some coastal and offshore planktonic algae. *Ecology* **42**, 444–446 (1961).
- Sunda, W. G., Swift, D. G. & Huntsman, S. A. Low iron requirements for growth in oceanic phytoplankton. *Nature* **351**, 55–57 (1991).
- Maldonado, M. T. & Price, N. M. Influence of N substrate on Fe requirements of marine centric diatoms. *Mar. Ecol. Prog. Ser.* **141**, 161–172 (1996).
- Peers, G., Quesnel, S.-A. & Price, N. M. Copper requirements for iron acquisition and growth of coastal and oceanic diatoms. *Limnol. Oceanogr.* **50**, 1149–1158 (2005).
- Raven, J. A., Evans, M. C. W. & Korb, R. E. The role of trace metals in photosynthetic electron transport in O₂-evolving organisms. *Photosynth. Res.* **60**, 111–149 (1999).
- Brand, L. E., Sunda, W. G. & Guillard, R. R. L. Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron. *Limnol. Oceanogr.* **28**, 1182–1198 (1983).
- Bruland, K. W. Oceanographic distributions of cadmium, zinc, nickel, and copper in the North Pacific. *Earth Planet. Sci. Lett.* **47**, 176–198 (1980).
- Moffett, J. W., Brand, L. E., Croot, P. L. & Barbeau, K. A. Cu speciation and cyanobacterial distribution in harbors subject to anthropogenic Cu inputs. *Limnol. Oceanogr.* **42**, 789–799 (1997).
- Tortell, P. D. & Price, N. M. Cadmium toxicity and zinc limitation in centric diatoms of the genus *Thalassiosira*. *Mar. Ecol. Prog. Ser.* **138**, 245–254 (1996).
- Sunda, W. G. & Huntsman, S. A. Regulation of copper concentrations in the oceanic nutricline by phytoplankton uptake and regeneration cycles. *Limnol. Oceanogr.* **40**, 132–137 (1995).
- Inda, L. A., Erdner, D. L., Peleato, M. L. & Anderson, D. M. Cytochrome *c*₆ isolated from the marine diatom *Thalassiosira weissflogii*. *Phytochemistry* **51**, 1–4 (1999).
- Papageorgiou, G. C. & Govindjee (eds) *Chlorophyll a Fluorescence: A Signature of Photosynthesis* (Kluwer Academic, Dordrecht, 2004).
- Weigel, M. *et al.* Plastocyanin is indispensable for photosynthetic electron flow in *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 31286–31289 (2003).
- Quinn, J. M. & Merchant, S. Copper-responsive gene expression during adaptation to copper deficiency. *Methods Enzymol.* **297**, 263–279 (1998).
- Schottler, M. A., Kirchhoff, H. & Weis, E. The role of plastocyanin in the adjustment of the photosynthetic electron transport to the carbon metabolism in tobacco. *Plant Physiol.* **136**, 4265–4274 (2004).
- da Silva, J. J. R. F. & Williams, R. J. P. *The Biological Chemistry of the Elements* (Oxford Univ. Press, New York, 2001).

21. La Fontaine, S. *et al.* Copper-dependent iron assimilation pathway in the model photosynthetic eukaryote *Chlamydomonas reinhardtii*. *Eukaryot. Cell* **1**, 736–757 (2002).
22. Armbrust, E. V. & Galindo, H. M. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus *Thalassiosira*. *Appl. Environ. Microbiol.* **67**, 3501–3513 (2001).
23. Archibald, J. M., Rogers, M. B., Topp, M., Ishida, K. & Keeling, P. J. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc. Natl Acad. Sci. USA* **100**, 7678–7683 (2003).
24. Strzepek, R. F. *Photosynthetic Iron Requirements of Marine Diatoms*. Thesis, Univ. British Columbia (2003).
25. Moffett, J. W., Zika, R. G. & Brand, L. E. Distribution and potential sources and sinks of copper chelators in the Sargasso Sea. *Deep-Sea Res.* **37**, 27–36 (1990).
26. Coale, K. H. & Bruland, K. W. Copper complexation in the Northeast Pacific. *Limnol. Oceanogr.* **33**, 1084–1101 (1988).
27. Bruland, K. W., Donat, J. R. & Hutchins, D. A. Interactive influences of bioactive trace metals on biological production in oceanic waters. *Limnol. Oceanogr.* **36**, 1555–1577 (1991).
28. Peers, G. & Price, N. M. A role for manganese in superoxide dismutases and growth of Fe-deficient diatoms. *Limnol. Oceanogr.* **49**, 1774–1783 (2004).
29. Chang, S. I. & Reinfelder, J. R. Bioaccumulation, subcellular distribution, and trophic transfer of copper in a coastal marine diatom. *Environ. Sci. Technol.* **34**, 4931–4935 (2000).
30. Navarro, J. A., Hervas, M. & De la Rosa, M. A. in *Photosynthesis Research Protocols* (ed. Carpentier, R.) 79–93 (Humana, Totowa, New Jersey, 2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Popovic for use of the PEA and M. R. Di Falco for purifying the trypsin-digested peptides. This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Centre for Environmental Bio-Inorganic Chemistry (CEBIC, Princeton University).

Author Information The sequences reported here have been deposited in the UniProt knowledgebase under accession number P84800. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.P. (graham.peers@mail.mcgill.ca).

Horizontal endosymbiont transmission in hydrothermal vent tubeworms

Andrea D. Nussbaumer¹, Charles R. Fisher² & Monika Bright¹

Transmission of obligate bacterial symbionts between generations is vital for the survival of the host. Although the larvae of certain hydrothermal vent tubeworms (Vestimentifera, Siboglinidae) are symbiont-free and possess a transient digestive system, these structures are lost during development, resulting in adult animals that are nutritionally dependent on their bacterial symbionts. Thus, each generation of tubeworms must be newly colonized with its specific symbiont^{1,2}. Here we present a model for tubeworm symbiont acquisition and the development of the symbiont-housing organ, the trophosome. Our data indicate that the bacterial symbionts colonize the developing tube of the settled larvae and enter the host through the skin, a process that continues through the early juvenile stages during which the trophosome is established from mesodermal tissue. In later juvenile stages we observed massive apoptosis of host epidermis, muscles and undifferentiated mesodermal tissue, which was coincident with the cessation of the colonization process. Characterizing the symbiont transmission process in this finely tuned mutualistic symbiosis provides another model of symbiont acquisition and additional insights into underlying mechanisms common to both pathogenic infections and beneficial host–symbiont interactions.

Symbiosis has had numerous central roles in the evolution of eukaryotic life, as well as in the evolution of life itself. Understanding the basic mechanisms of symbiosis is crucial to an appreciation of this widespread biological phenomenon. Symbiont transmittal between generations is key to the persistence of all symbiotic associations. In many well-studied cases, including a variety of nitrogen-fixing³, bioluminescent⁴ and algal–invertebrate⁵ symbioses, transmittal between generations is indirect and each generation must be infected *de novo*. Horizontal symbiont transfer has been also suggested for the obligate vestimentiferan (Siboglinidae, Polychaeta) endosymbiosis^{1,2,6–9}. These tubeworms thrive at deep-sea hydrothermal vents and cold seeps rich in reduced chemicals, where they often dominate the community^{10,11}. The animals lack a digestive system as adults^{12,13} and their nutritional needs are met by their bacterial chemoautotrophic symbionts contained in a morphologically complex symbiont-housing organ called the trophosome^{8,12–14}.

The larvae of vestimentiferans disperse over long distances to find new and ephemeral hydrothermal vents that are heavily colonized very quickly¹⁰. However, larvae are aposymbiotic and undergo an infection process and marked developmental changes on settlement. The currently accepted hypothesis^{6,7} suggests that symbionts enter through the larval mouth and digestive system and that the subsequent proliferation of endodermal midgut cells leads to the development of the trophosome. This was based on the presence of a functional mouth and a gut containing microbes in early post-larval tubeworms^{7,8}. Here we present data that indicate a very different mode of symbiont acquisition and development of the trophosome. Although we did not follow the infection and developmental processes in live

animals directly, our ultrastructural and developmental data from a series of animals of different sizes with symbionts localized by using molecular methods are not consistent with symbiont uptake through the post-larval mouth or with an endodermal developmental origin of the trophosome. We propose that symbionts infect the skin of the larvae after settlement, migrate to the dorsal mesentery and proliferate to form the trophosome, while the infection process ceases simultaneously with extensive apoptosis of host cells in the early juvenile stage.

To sample newly settled larvae and very small juveniles we developed special settlement devices, namely tubeworm artificial settlement cubes (TASCs). TASCs, which are composed of individual grooved plates that can be taken apart, allowed us to locate and recover the very small and almost transparent tubeworm specimens alive and undamaged (Fig. 1). The three vestimentiferan species co-occurring at the hydrothermal vents of the 9° 50' N region of the East Pacific Rise—*Riftia pachyptila*, *Tevnia jerichonana* and *Oasisia alvinae*, which share virtually the same symbiotic phylotype^{2,9}—were established on the TASCs after one year of deployment. Although larger animals of these three species between about 2 mm and 40 cm length could be identified (Fig. 1a), small tubeworm larvae and juveniles (200 µm to about 2 mm) are morphologically identical (Fig. 1b).

Laser-scanning confocal microscopy was used to localize the signal of the 16S or 23S ribosomal-RNA-labelled probes that were specific for the vent tubeworm symbiont and different groups of Bacteria and

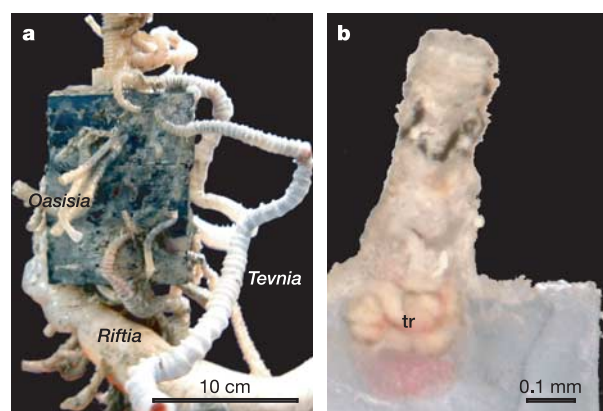


Figure 1 | Tubeworm artificial settlement cubes (TASCs). **a**, *Riftia pachyptila*, *Oasisia alvinae* and *Tevnia jerichonana* on TASCs after one year of deployment at East Pacific Rise, site TICA (9° 50.447' N, 104° 17.493' W, 2,500 m depth), recovered in December 2002. **b**, Close-up of one settled juvenile tubeworm 400 µm in length, with trophosome (tr) and tube (700 µm in length).

¹Department of Marine Biology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria. ²Department of Biology, Pennsylvania State University, 208 Mueller Laboratory, University Park, Pennsylvania 16802, USA.

Archaea, on series of semithin sections (Supplementary Table 1). With the use of detailed reconstructions of serial transmission electron microscopy (t.e.m.) ultrathin sections (Fig. 2a, b, and Supplementary Fig. 1) and fluorescence *in situ* hybridization (FISH; Fig. 2c), we confirmed that vestimentiferan larvae are aposymbiotic with a fully developed transient digestive tract on settlement. We examined several segmented trochophore larvae (200–250 μm in length) (Supplementary Fig. 1). These animals had a larval locomotory ciliary girdle (prototoch), a ventrally located patch of cilia (neurotroch) and two or three coelomic body cavities with fully differentiated mesoderm tissue¹⁵. The development of some adult characters, such as the first pair of tentacles of the gas exchange organ (plume), the tube-building pyriform glands and hard structures (chaetae) to maintain the position of the animal in its tube, had begun in these larvae. The smallest larvae examined lacked a trophosome, and no symbionts were detected in any tissue or the gut lumen (Fig. 2c). The digestive tract was transient and was composed of a ventral mouth opening, a short buccal cavity, three regions, distinguished by microanatomical features as foregut, midgut and hindgut (length ratio 4:5:1), and a terminally located anus. Whereas the foregut and the hindgut had similar, numerous electron-dense granules and rough endoplasmic reticulum, the midgut was distinctly different and characterized by large glandular vesicles. A few bacteria, but no symbionts, and tests of protists were occasionally detected in the gut lumen as well as in gut cells, where they were apparently undergoing degradation (Fig. 2d, e). Thus, in contrast to laboratory-reared embryos of several hydrothermal vent and cold seep vestimentiferans that developed to an early non-feeding trochophore larval stage lacking a mouth^{16,17}, the newly settled later-stage segmented trochophore were apparently actively feeding on microbes.

Infection with symbionts was first apparent in larvae about 250 μm in length in the series of ultrathin and semithin sections viewed in transmission electron and light microscopy (Figs 2a, b and 3a, and Supplementary Fig. 2a) and by FISH performed on serial semithin sections (Fig. 3c, and Supplementary Fig. 2a). The newly infected larvae were virtually identical in morphology and size to the aposymbiotic specimens examined from the same collections, further supporting the suggestion that this is the stage at which infection begins.

The skin was identified as the sole site of symbiont entrance into the host. Fewer than 20 symbionts were found to infect the epidermis, adjoining muscles and underlying undifferentiated mesodermal tissue in the larvae examined (Fig. 3d). We estimated that at least fivefold more symbionts (more than 100) were present in small juveniles up to 400 μm in length but the infected tissues were the same (Supplementary Fig. 2b). Rod-shaped bacteria of the symbiotic phylotype were present in the dorsal mesentery, the mesodermal region between the dorsal blood vessel and the foregut (Fig. 3a–c, and Supplementary Fig. 2a). The few symbionts enclosed in vacuoles of mesodermal cells apparently initiate the development of the trophosome and thus represent the very beginning of the symbiosis in vestimentiferan tubeworms. When the symbionts occurred in the epidermis, muscles and undifferentiated mesoderm, they were dispersed freely within the cytoplasm (Fig. 3e) and between these cells. However, in the trophosome, the symbionts were always enclosed in vacuoles of host cells (Fig. 3b). We did not detect symbionts in the gut lumen or gut cells, the mouth opening or the anus in any infected larvae. Apoptosis of both the symbiont-infected and adjacent non-infected cells in the epidermis, muscles and undifferentiated mesoderm was demonstrated in small juveniles, but not in larvae, by cell shrinkage, pycnotic nuclei with dense chromatin masses accumulating peripherally, and dilated mitochondria^{18,19} (Fig. 3e, and Supplementary Fig. 3). In larger juveniles and adults the symbionts were present only in the trophosome (Supplementary Fig. 4).

We used the same probes and FISH as well as t.e.m. to examine the microbial population associated with the developing tube in larvae

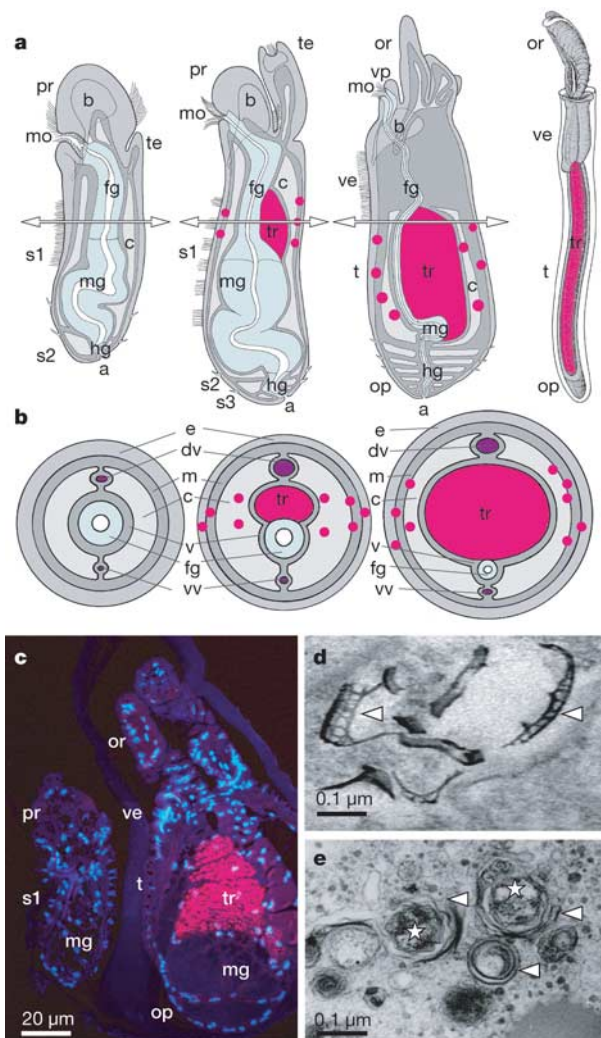


Figure 2 | Symbiont acquisition and early development of recently settled vestimentiferans. Schematic sagittal drawings of animals reconstructed from serial sections (a), and schematic cross-sections in the region of symbiont uptake and trophosome development (outlined with double arrow in sagittal drawing) (b), fluorescence *in situ* hybridization (FISH) micrograph (c) and t.e.m. micrographs of larval midgut (d, e). **a**, Left: aposymbiotic larva with prostomium (pr) containing brain (b) and peristomium with mouth opening (mo), two segments (s1, s2) with corresponding coelomic cavities (light grey) (c) and tentacles (te) developing from the first segment. Left centre: infected larva with three segments (s1–s3); symbionts (pink) in the skin and the dorsal mesentery where the trophosome (tr, pink) starts to develop. Right centre: juvenile with prostomium, peristomium (transformed to ventral process (vp) with mouth opening) and anterior part of first segment merged to the vestimentum (ve), obturacular region (or) developing from vestimentum, posterior part of the first segment elongated to trunk (t) and posterior located segments merged to opisthosome (op); at this stage symbionts and apoptosis in the skin co-occur. Right: adult body regions are similar to those in juveniles. A transient digestive tract (blue) composed of mouth opening (mo), foregut (fg), midgut (mg), hindgut (hg) and anus (a) is present in larvae and small juveniles but is reduced in later stages. **b**, Corresponding cross-sections of aposymbiotic larva with foregut (fg, blue) surrounded by visceral mesoderm (v) and dorsal (dv) and ventral (vv) blood vessels, coelom (c) and skin with epidermis (e) and muscles (m); in addition, symbionts (pink) in the skin and the trophosome (tr, pink) are present in infected larva and juvenile. **c**, FISH on LR White sections with symbiont-specific probes labelled with Cy3 (red), 4',6-diamidino-2-phenylindole (DAPI) counterstain (blue): aposymbiotic larva (left), juvenile with labelled symbionts confined to trophosome (right); abbreviations as in **a**. **d**, T.e.m. micrograph of aposymbiotic larva showing midgut cells with degrading protists and remaining tests (arrowheads). **e**, Cells as in **d**, containing degrading bacteria (stars) surrounded by myelin bodies (arrowheads).

and small juveniles and the established solid tubes in larger juveniles and adults. In larvae, an extracellular substance apparently released by pyriform glands formed a mucous coat in which newly settled animals were embedded. A microbial assemblage of diverse environmental bacteria that included the symbionts colonized this mucous coat (Fig. 3f, and Supplementary Fig. 5a, b). We never detected the symbiotic phylotype in established tubes of larger juveniles and adults, although Gammaproteobacteria, Alphaproteobacteria and Bacteroidetes (*Cytophaga-Flavobacter*) were present (Supplementary Fig. 5c–f).

The previously accepted hypothesis for symbiont acquisition and trophosome development in vestimentiferans suggested that symbionts enter through the larval mouth with other food and evade

digestion in the gut, and the trophosome develops from proliferating (endodermal) midgut^{6,7} tissue. We found no evidence for this process: the symbiont phylotype was never present in the digestive system, and the midgut tissues did not show any signs of proliferation or involvement with symbionts. Furthermore, the first presumptive bacteriocytes of the trophosome in infected larvae were located in the foregut region, in mesodermal cells of the dorsal mesentery between the dorsal blood vessel and the foregut. The most parsimonious interpretation of our data indicates a different developmental pathway for symbiont acquisition and development of the trophosome: symbionts are taken up across epidermal cells during a symbiont-specific selective infection process and subsequently migrate through several layers of host tissue into a mesodermal tissue¹⁴ that will develop into the trophosome. Once the trophosome is well established in juveniles, the infection process ceases at the same time as massive apoptosis of skin and other non-trophosome symbiont-containing tissues.

Apoptosis during ontogenesis²⁰, tissue renewal²⁰ and pathogenic infections²¹ is well documented. During infections of *Salmonella* ssp. and *Shigella* ssp. the initiation of inflammation and the destruction of host tissue are mediated by apoptosis caused by the pathogens; however, host apoptosis as a defence mechanism against *Mycobacterium bovis* BCG is also known^{21,22}. The involvement of apoptosis in symbiont acquisition by other mutualistic symbioses has also been shown. In the bioluminescent *Vibrio*–squid symbiosis, the infection process is facilitated through pores of a ciliated epithelium located on the surface of the host. As soon as the symbionts have colonized the developing light organ, the epithelium degenerates through apoptosis; uptake then ceases²³. We suggest a similar developmental process, because we found massive apoptosis in the tissues mediating the symbiont uptake—the skin—after the symbionts have become well established in the target organ, the trophosome. Now that the principal mode of acquisition and developmental processes in the vestimentiferan tubeworm symbiosis are better understood, specific underlying mechanisms that control these processes can be addressed.

METHODS

Collection of vestimentiferans. TASCs consisted of ten HPVC plates (5 cm × 5 cm × 0.5 cm) with 20 grooves (20 mm × 1 mm × 1 mm) on one side of each, bolted together. One assembled TASC offers 200 holes providing some physical protection for juvenile tubeworms. TASCs were deployed and recovered by DSV *Alvin* at the hydrothermal vent site TICA (9° 50.447' N, 104° 17.493' W; 2,500 m depth) in December 2001, 2002 and 2003. On recovery, TASCs were disassembled so that the fragile and translucent specimens could be removed under a stereomicroscope and fixed appropriately for further study. Some tubeworms were also collected from tubes of adult tubeworms, ambient basalt pieces or artificial basalt blocks²⁴ from different vent sites along the 9° 50' N region of the East Pacific Rise in 1998, 1999 and 2001.

Specimen fixation and preparation. For FISH, larvae and juveniles were fixed in 4% paraformaldehyde, 0.1 M PBS pH 7.4 containing 10% (w/v) sucrose at 4 °C for 12 h and embedded in paraffin or in medium-grade LR White resin (British BioCell International). Complete series of 4-μm paraffin sections or alternating 1-μm and 70-nm resin sections were cut. For t.e.m., one larva and three juveniles were fixed, embedded in Spurr resin and serial-sectioned as described previously¹⁴. Schematic drawings of one aposymbiotic larva, one symbiotic larva and one juvenile were constructed with t.e.m. images of at least every 3 μm. For detailed protocols see Supplementary Methods.

Fluorescent *in situ* hybridization (FISH). FISH was performed as described previously²⁵, with modifications for the LR White sections (Supplementary Methods). Hybridizations were performed with three newly designed Cy3-labelled oligonucleotide probes (from Thermo electron and GenExpress) specific for the symbiont 16S rRNA of *Riftia pachyptila*, *Tevnia jerichonana* and *Oasisia alviniae* (Rif/Tev/Oas; Supplementary Table 1). The probes were designed with the respective tool of the ARB software package²⁶ and checked for similarity by using the Probe Match tool integrated in ARB, against all 16S rRNA sequences in the Ribosomal Database Project²⁷, and by using GenBank²⁸.

The optimal hybridization conditions for the Rif/Tev/Oas probes were determined with increasing formamide concentrations on adult *Riftia* trophosome embedded in paraffin, in LR White and on isolated symbiont preparations

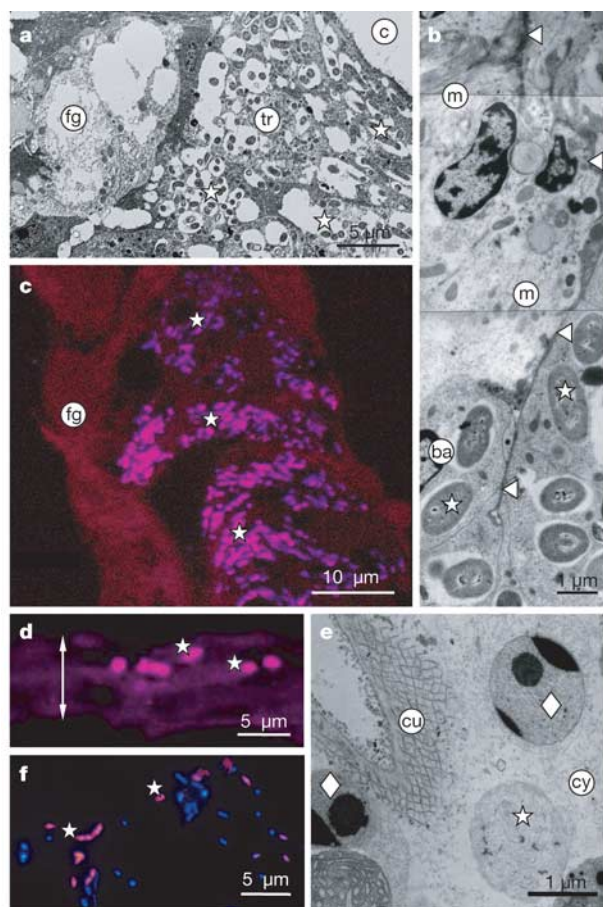


Figure 3 | Infection and trophosome development in vestimentiferans. **a**, T.e.m. micrograph of larva; rod-shaped symbionts (stars) initiating the development of the trophosome (tr) in the dorsal mesentery region next to the foregut (fg) and the coelom (c). **b**, T.e.m. micrograph of epidermis in the dorsal mesentery with epithelio-muscle cells (m) next to peritoneal cells, now called bacteriocytes (ba) with rod-shaped symbionts (stars) enclosed in vacuoles; both epithelial cell types share a continuous basal matrix (arrowheads). This organization supports the mesodermal origin of the bacteriocytes. **c**, Same specimen as in **a**: laser scanning confocal micrograph (LSCM) of FISH; overlay of two images with symbiont-specific probes (red) and universal eubacterial probe mix (blue) showing the rod-shaped symbionts (stars) initiating the trophosome development next to the foregut (fg) and the absence of symbionts in the gut. **d**, LSCM of FISH with the same probes as in **c**: specific infection with rod-shaped symbionts (stars) in the skin (outlined with double arrow). **e**, T.e.m. micrograph of epidermis from small juvenile with symbiont-like bacteria (stars) in the cytoplasm (cy), apoptotic nuclei (diamonds) and cuticle (cu). **f**, LSCM of FISH of extracellular substance located on the cuticle of a segmented larva shows a microbial assemblage including the symbionts (stars); symbiont-specific probes labelled with Cy3 (red) and DAPI counterstain (blue).

until the probe signal was no longer detectable. For a positive control all sections were hybridized simultaneously with the symbiont-specific probes and the Bacteria probe mix²⁹ labelled with different dyes. Three negative controls were employed: first, symbiont-specific probes were tested on other Gammaproteobacteria chemoautotrophic symbionts (from the ciliate *Zoothamnium niveum* and the nematode *Laxus cosmopolitus*); second, a version of the symbiont-specific probe with one centrally located mismatch was tested at the respective stringency on *Riftia* trophosome tissue; and third, a nonsense probe (NON 338)³⁰ was tested on one section (out of four) on every slide investigated.

Received 20 December 2005; accepted 10 April 2006.

- Cary, S. C., Warren, W., Anderson, E. & Giovannoni, S. Identification and localization of bacterial endosymbionts in hydrothermal vent taxa with symbiont-specific polymerase chain reaction amplification and in situ hybridization techniques. *Mol. Mar. Biol. Biotechnol.* **2**, 51–62 (1993).
- Di Meo, C. A. et al. Genetic variation among endosymbionts of widely distributed vestimentiferan tubeworms. *Appl. Environ. Microbiol.* **66**, 651–658 (2000).
- van Rhijn, P. & Vanderleyden, J. The *Rhizobium*–plant symbiosis. *Microbiol. Rev.* **59**, 124–142 (1995).
- McFall-Ngai, M. J. & Ruby, E. G. Developmental biology in marine invertebrate symbioses. *Curr. Opin. Microbiol.* **3**, 603–607 (2000).
- Trench, R. K. Microalgal–invertebrate symbioses: a review. *Endocytobiosis Cell Res.* **9**, 135–175 (1993).
- Jones, M. L. & Gardiner, S. L. Evidence for a transient digestive tract in Vestimentifera. *Proc. Biol. Soc. Wash.* **101**, 423–433 (1988).
- Southward, E. C. Development of the gut and segmentation of newly settled stages of *Ridgeia* (Vestimentifera): Implications for relationship between Vestimentifera and Pogonophora. *J. Mar. Biol. Ass. UK* **68**, 465–487 (1988).
- Callsen-Cencic, P. & Flügel, H. J. Larval development and the formation of the gut *Siboglinum poseidonii* Flügel & Langhof (Pogonophora, Perviatia), evidence of protostomian affinity. *Sarsia* **80**, 73–89 (1995).
- Feldman, R. A., Black, M. B., Cary, C. S., Lutz, R. A. & Vrijenhoek, R. C. Molecular phylogenetics of bacterial endosymbionts and their vestimentiferan hosts. *Mol. Mar. Biol. Biotechnol.* **6**, 268–277 (1997).
- Shank, T. M. et al. Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'N, East Pacific Rise). *Deep-sea Res. II* **45**, 465–515 (1998).
- Cordes, E. E., Arthur, M. A., Shea, K., Arvidson, R. S. & Fisher, C. R. Modeling the mutualistic interactions between tubeworms and microbial consortia. *PLoS Biol.* **3**, e77 (2005).
- Rouse, G. W., Goffredi, S. K. & Vrijenhoek, R. C. *Osedax*: bone-eating marine worms with dwarf males. *Science* **305**, 668–671 (2004).
- Southward, E. C., Schulze, A. & Gardiner, S. L. Pogonophora (Annelida): form and function. *Hydrobiologia* **535/536**, 227–251 (2005).
- Bright, M. & Sorgo, A. Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebr. Biol.* **122**, 345–366 (2003).
- Heimler, H. Larvae. *Microfauna Marina* **4**, 353–371 (1988).
- Young, C. M., Vázquez, E., Metaxas, A. & Tyler, P. A. Embryology of vestimentiferan tube worms from deep-sea methane/sulphide seeps. *Nature* **381**, 514–516 (1996).
- Marsh, A. G., Mullineaux, L. S., Young, C. M. & Manahan, D. T. Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents. *Nature* **411**, 77–80 (2001).
- Vaux, D. Toward an understanding of the molecular mechanisms of physiological cell death. *Proc. Natl Acad. Sci. USA* **90**, 786–789 (1993).
- Clarke, P. G. Developmental cell death: morphological diversity and multiple mechanisms. *Anat. Embryol. (Berl.)* **181**, 195–213 (1990).
- Twomey, C. & McCarthy, J. V. Pathways of apoptosis and importance in development. *J. Cell. Mol. Med.* **9**, 345–359 (2005).
- Zychlinsky, A. & Sansonetti, P. Apoptosis in bacterial pathogenesis. *J. Clin. Invest.* **100**, 493–495 (1997).
- Molloy, A., Laochumroonvorapong, P. & Kaplan, G. Apoptosis, but not necrosis, of infected monocytes is coupled with killing of intracellular bacillus *Calmette-Guerin*. *J. Exp. Med.* **180**, 1499–1509 (1994).
- Foster, J. S. & McFall-Ngai, M. J. Induction of apoptosis by cooperative bacteria in the morphogenesis of host epithelial tissues. *Dev. Genes Evol.* **208**, 295–303 (1998).
- Mullineaux, L. S., Fisher, C. R., Peterson, C. H. & Schaeffer, S. W. Tubeworm succession at hydrothermal vents: use of biogenic cues to reduce habitat selection error? *Oecologia* **123**, 275–284 (2000).
- Dubilier, N., Giere, O., Distel, D. L. & Cavanaugh, C. M. Characterization of chemoautotrophic symbionts in a gutless marine worm (Oligochaeta, Annelida) by phylogenetic 16S rRNA sequence analysis and in situ hybridization. *Appl. Environ. Microbiol.* **61**, 2346–2350 (1995).
- Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
- Cole, J. R. et al. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33**, D294–D296 (2005).
- Benson, D. A. et al. GenBank. *Nucleic Acids Res.* **27**, 12–17 (1999).
- Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: Development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).
- Manz, W., Amann, R., Ludwig, W., Wagner, M. & Schleifer, K. H. Phylogenetic oligonucleotide probes for the major subclasses of Proteobacteria: Problems and solutions. *Syst. Appl. Microbiol.* **15**, 593–600 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the captain and crew of the RV *Atlantis* and the crew of the DSV *Alvin* for their continuous support; H. Grillitsch for the schematic illustrations; P. Gahleitner for sectioning; W. Klepal for the EM support; S. C. Cary, C. M. Cavanaugh (grants from NOAA and NSF), J. J. Childress and L. S. Mullineaux for their hospitality on cruises; and C. M. Cavanaugh and M. Horn for their comments on the manuscript. This work was supported by grants from the Austrian Science Foundation and the Austrian Academy of Science to M.B., by a grant from the Faculty of Life Sciences, University of Vienna to A.D.N., and a grant from the US National Science Foundation to C.R.F.

Author Contributions M.B. was the project leader, designed the TASCs and performed the t.e.m. work; A.D.N. performed the molecular work and was responsible for the data collection; and A.D.N., C.R.F. and M.B. performed the field work and wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (monika.bright@univie.ac.at).

Polycomb complexes repress developmental regulators in murine embryonic stem cells

Laurie A. Boyer^{1*}, Kathrin Plath^{1*†}, Julia Zeitlinger¹, Tobias Brambrink¹, Lea A. Medeiros^{1,2}, Tong Ihn Lee¹, Stuart S. Levine¹, Marius Wernig¹, Adriana Tajonar², Mridula K. Ray², George W. Bell¹, Arie P. Otte³, Miguel Vidal⁴, David K. Gifford⁵, Richard A. Young^{1,2} & Rudolf Jaenisch^{1,2}

The mechanisms by which embryonic stem (ES) cells self-renew while maintaining the ability to differentiate into virtually all adult cell types are not well understood. Polycomb group (PcG) proteins are transcriptional repressors that help to maintain cellular identity during metazoan development by epigenetic modification of chromatin structure¹. PcG proteins have essential roles in early embryonic development^{2–6} and have been implicated in ES cell pluripotency⁷, but few of their target genes are known in mammals. Here we show that PcG proteins directly repress a large cohort of developmental regulators in murine ES cells, the expression of which would otherwise promote differentiation. Using genome-wide location analysis in murine ES cells, we found that the Polycomb repressive complexes PRC1 and PRC2 co-occupied 512 genes, many of which encode transcription factors with important roles in development. All of the co-occupied genes contained modified nucleosomes (trimethylated Lys 27 on histone H3). Consistent with a causal role in gene silencing in ES cells, PcG target genes were de-repressed in cells deficient for the PRC2 component Eed, and were preferentially activated on induction of differentiation. Our results indicate that dynamic repression of developmental pathways by Polycomb complexes may be required for maintaining ES cell pluripotency and plasticity during embryonic development.

Biochemical and genetic evidence indicates that PcG proteins function in two distinct complexes, PRC1 and PRC2, the core components of which are conserved from fruitfly to human and are essential for PcG activity both *in vitro* and *in vivo*¹. To gain insights into the role of PcG proteins in ES cells, we identified the genes occupied by PcG proteins in murine ES cells by performing genome-wide location analysis using antibodies against core components of PRC1 (Phc1 and Rnf2) and PRC2 (Suz12 and Eed) (Fig. 1 and Supplementary Fig. S1).

The genomic DNA associated with PcG proteins and total (control) DNA were combined and hybridized to microarrays that contained 60-mer oligonucleotide probes covering the region from –8 kb to +2 kb relative to the transcription start sites for 15,742 annotated mouse genes (see Supplementary Information). Genomic sites occupied by PcG proteins (Supplementary Tables S1–S4) were identified as peaks of chromatin immunoprecipitation (ChIP)-enriched DNA using a previously validated algorithm⁷ (Fig. 1a and Supplementary Information). Notably, the vast majority (~90%, $P < 10^{-159}$) of bound sequences were detected within 1 kb of a transcription start site (Supplementary Fig. S2). We focused further

analysis on this set of bound regions, as the binding of a regulator in close proximity to a transcription start site is probably associated with the regulation of that gene. Our analysis revealed that PRC1 and PRC2 components occupied an overlapping set of target genes, and identified with high confidence 512 genes bound by all four PcG proteins in ES cells (Fig. 1b, Supplementary Tables S5 and S6, and Supplementary Figs S3–S5). These data show that PRC1 and PRC2 co-occupy the promoter regions of a large set of genes in ES cells.

Because trimethylation of Lys 27 on histone H3 (H3K27me3) is thought to be a marker for repressive chromatin and is associated with PRC2 activity^{8–12}, we investigated whether PcG occupancy correlated with this histone modification across the genome. We found that H3K27me3 was enriched at all 512 genes occupied by components of both PRC1 and PRC2 complexes (Supplementary Tables S5 and S7). Analysis of the distribution of H3K27me3 at this set of genes revealed that, similar to PRC1 and PRC2 components, H3K27me3 was also associated with probes close to the transcription start site of PcG target genes (Fig. 1c). These data show that PRC1 and PRC2 occupancy is associated with modified nucleosomes, consistent with a repressed transcriptional state.

To gain insights into the biological role of PcG proteins in ES cell pluripotency, we determined which gene ontology (GO) terms were over-represented in the set of genes associated with both Polycomb complexes and H3K27me3 (Fig. 1d and Supplementary Table S8). This analysis revealed an extremely significant enrichment for genes connected to transcription and development hierarchies, including organogenesis, morphogenesis, pattern specification, neurogenesis, cell differentiation, embryonic development and cell-fate commitment, among others (Supplementary Fig. S6). Further analysis showed that the target genes within the development and transcription functional groups overlap significantly ($P < 10^{-96}$; Supplementary Table S8), indicating that most of the PcG target genes are transcription factors with important roles in a variety of developmental processes.

In addition to *Hox* genes, which are the classic PcG target genes, PcG proteins bound to over 100 genes encoding homeodomain-containing transcriptional regulators in murine ES cells, including members of the *Dlx*, *Irx*, *Lhx*, *Pou*, *Pax* and *Six* gene families (Supplementary Table S9 and Supplementary Fig. S7). Homeodomain-containing transcription factors are evolutionarily conserved regulators that specify cell fate during embryonic development through transcriptional control of other developmental regulators¹³. Polycomb complexes also occupied promoters of members of the

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 SM Amsterdam, The Netherlands. ⁴Developmental and Cell Biology Centro de Investigaciones Biológicas (CSIC), 28040 Madrid, Spain. ⁵Computer Science and Artificial Intelligence Laboratories, 32 Vassar Street, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: Biological Chemistry Department, David Geffen School of Medicine at UCLA, 36-133 CHS, PO Box 951737, Los Angeles, California 90095-1737, USA.

*These authors contributed equally to this work.

Fox, *Sox*, *Gata* and *Tbx* transcription factor families, which also have essential roles in development and disease^{14–17}. These data demonstrate that in ES cells, Polycomb complexes target transcription factors that have key roles in a variety of developmental processes.

To establish a direct functional link between PcG protein binding and repression of target genes, we used real-time polymerase chain reaction (PCR) to compare the expression levels of PcG target genes in wild-type ES cells and ES cells lacking the PRC2 component *Eed* (Fig. 2). We examined *Eed*-deficient cells (*Eed*^{−/−}) because lack of *Eed* leads to disruption of PRC2 and H3K27 methylation¹⁸ (Supplementary Fig. S8). Transcript levels for the PcG target genes *Gata3*, *Gata4* and *Gata6* increased significantly in *Eed* mutant cells compared with wild-type ES cells, whereas transcript levels for *Gata1*, which was unbound in the ChIP assay, were unchanged (Fig. 2a). Immunostaining showed that Gata4 protein was not detected in wild-type ES cells, but it was abundant in *Eed* mutant cells, consistent with our PCR results (Fig. 2b). Loss of *Eed* resulted in at least a

twofold increase in transcript abundance for 87% (71/82) of the PcG target genes tested, including members of the *Hox*, *Pax*, *Lhx*, *Sox*, *Fox* and *Fgf* gene families, whereas the expression of control genes remained largely unaffected (Fig. 2c and Supplementary Table S10). These data indicate that genes bound by PcG proteins in ES cells are direct targets of PRC2-mediated repression.

To investigate whether PRC1 promoter occupancy is dependent on PRC2 at target genes in ES cells, we performed ChIP experiments and site-specific PCR for a subset of genes in wild-type and *Eed* mutant cells (Fig. 2d). Previous work has shown that loss of PRC2 components results in disruption of PRC1 binding at *Hox* genes in somatic cells^{8,19}. In *Eed* mutant cells, we observed a significant reduction in binding at PcG target genes for two PRC1 components, Rnf2 and Cbx2, as well as the PRC2 subunit Suz12, even though the levels of these proteins had not dramatically changed in the absence of *Eed* (Supplementary Fig. S8). Our results suggest that an intact PRC2 complex and/or its associated histone methyl mark (H3K27me3) are required for the association of PRC1 at target genes in ES cells.

The association of PcG components with repressive chromatin structure and developmental regulators suggests that genes targeted by Polycomb complexes are globally repressed in ES cells but must be activated during differentiation. To test this, we analysed the expression of PcG target genes during ES cell differentiation (Fig. 3), and found that most PcG target gene transcripts were lower in abundance in ES cells relative to differentiated cells (Fig. 3a and Supplementary Table S11). Specifically, 93% of PcG target transcripts were upregulated during ES cell differentiation, compared with only 59% of all transcripts represented on the array (Fig. 3b). Further analysis using the Gene Set Enrichment Analysis (GSEA) tool (see Supplementary Information), which tests for non-random distribution of a subset of genes within a ranked expression data set, confirmed our hypothesis that PcG target genes are preferentially upregulated during ES cell differentiation (Fig. 3c). This suggests that PcG proteins have specialized roles in silencing genes in ES cells, the activation of which correlates with differentiation and loss of pluripotency.

These findings suggest that PRC2 binding and H3K27 methylation should be decreased at target genes that are activated during lineage commitment. To test this directly, we analysed expression levels and H3K27me3 status for a subset of genes during directed differentiation of ES cells to neural precursor (NP) cells using real-time PCR upon reverse transcription and ChIP with site-specific real-time PCR, respectively, as we noted that many PcG target genes have known roles in neural development (Table 1 and Supplementary Fig. S9). We found that H3K27me3 was depleted at least 100-fold at neural-specific genes (for example, *Olig1*, *Olig2*, *Olig3* and *Nes*), and that expression of these genes was significantly increased in NP cells relative to ES cells (Table 1, group A). The loss of H3K27me3 was concomitant with RNA polymerase II occupancy and an increase in H3K4me3, the histone methylation mark associated with active transcription (data not shown). In contrast, PcG target genes that were not expressed in either ES or NP cells retained high levels of H3K27me3 in their promoter regions and were not associated with RNA polymerase II (Table 1, group B and data not shown). We also found that the pluripotency genes *Oct4* (also known as *Pou5f1*) and *Nanog* were associated with low levels of H3K27me3 in ES cells, consistent with their high expression (Table 1, group C). Although these genes become silenced on ES cell differentiation, H3K27me3 levels were similar in ES and NP cells, consistent with previous studies suggesting that *Oct4* is repressed by other epigenetic silencing mechanisms^{20–22}. Together, these data demonstrate that genes repressed by PcG proteins in ES cells maintain the potential to become activated on lineage commitment, revealing a dynamic role for PcG complexes and their chromatin modifications during differentiation.

This dynamic role for PcG-mediated gene repression seems to be

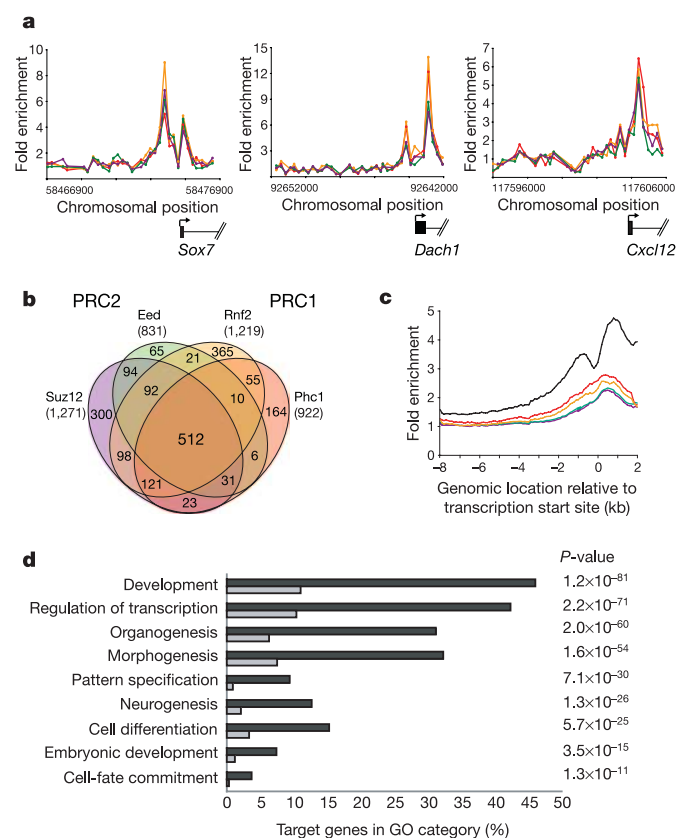


Figure 1 | PRC1 and PRC2 colocalize at genes encoding developmental regulators. **a**, Unprocessed enrichment ratios for all probes within a genomic region (ChIP-enriched versus total genomic DNA) for Suz12 (green), Eed (purple), Rnf2 (orange) and Phc1 (red). Example genes are drawn to scale below plots, and the start and direction of transcription are noted by arrows. **b**, Venn diagram showing the overlap among genes bound by PcG proteins within 1 kb of a transcription start site. Numbers in parentheses represent the total number of genes bound by the respective PcG protein. **c**, Average unprocessed enrichment ratios for each oligonucleotide probe within the −8 kb to +2 kb genomic region for all co-bound RefSeq genes relative to the transcription start site. Colours represent PcG proteins as in **a**, and the black line shows H3K27me3 enrichment. **d**, Gene Ontology analysis of PcG target genes. Black bars represent the observed percentage of PcG target genes in a particular GO category. Grey bars represent the percentage expected on the basis of all GO-annotated genes on the oligonucleotide array. The significance (*P*-value) of this enrichment is based on a hypergeometric distribution.

different from other epigenetic silencing mechanisms. For example, DNA methylation is thought to be a stable silencing mechanism that is required for irreversibly locking-in the repressed transcriptional state. It has recently been shown that PRC2 recruits DNA methylation machinery to silence target genes in somatic cells²³. Our finding that PcG proteins repress genes that are poised for activation on differentiation suggests that PcG-mediated repression is not linked to DNA methylation in ES cells. Consistent with this idea, DNA methylation does not have an essential role in maintaining ES cells in an undifferentiated state, but is essential for the survival of somatic cells^{24,25}. In contrast, ES cells cannot be derived from blastocysts deficient for the PRC2 component *Ezh2* (ref. 2), and ES cells lacking *Eed* have a strong propensity to differentiate (Supplementary Fig. S8), indicating that PcG-mediated gene repression in ES cells and during early development may be important for developmental plasticity.

PcG proteins are recruited to target sites through interaction with site-specific DNA-binding proteins in *Drosophila*¹. It is currently unknown how PcG proteins are specifically recruited to target genes in mammals. Many of the developmental PcG target genes identified in this study in murine ES cells have recently been shown to be bound by three key pluripotency transcription factors—*OCT4* (*POU5F1*), *SOX2* and *NANOG*—in human ES cells⁷ (see Supplementary Information). Moreover, it has recently been shown that PcG proteins target a similar set of developmental regulators in human ES cells²⁶. This raises the possibility that PcG proteins, at least at a subset of genes, act as transcriptional repressors by collaborating with a specific set of transcription factors. Improved understanding of the mechanisms by which these transcription factors and PcG proteins contribute to maintenance of the pluripotent state and repression of developmental genes remain important issues for future investigation.

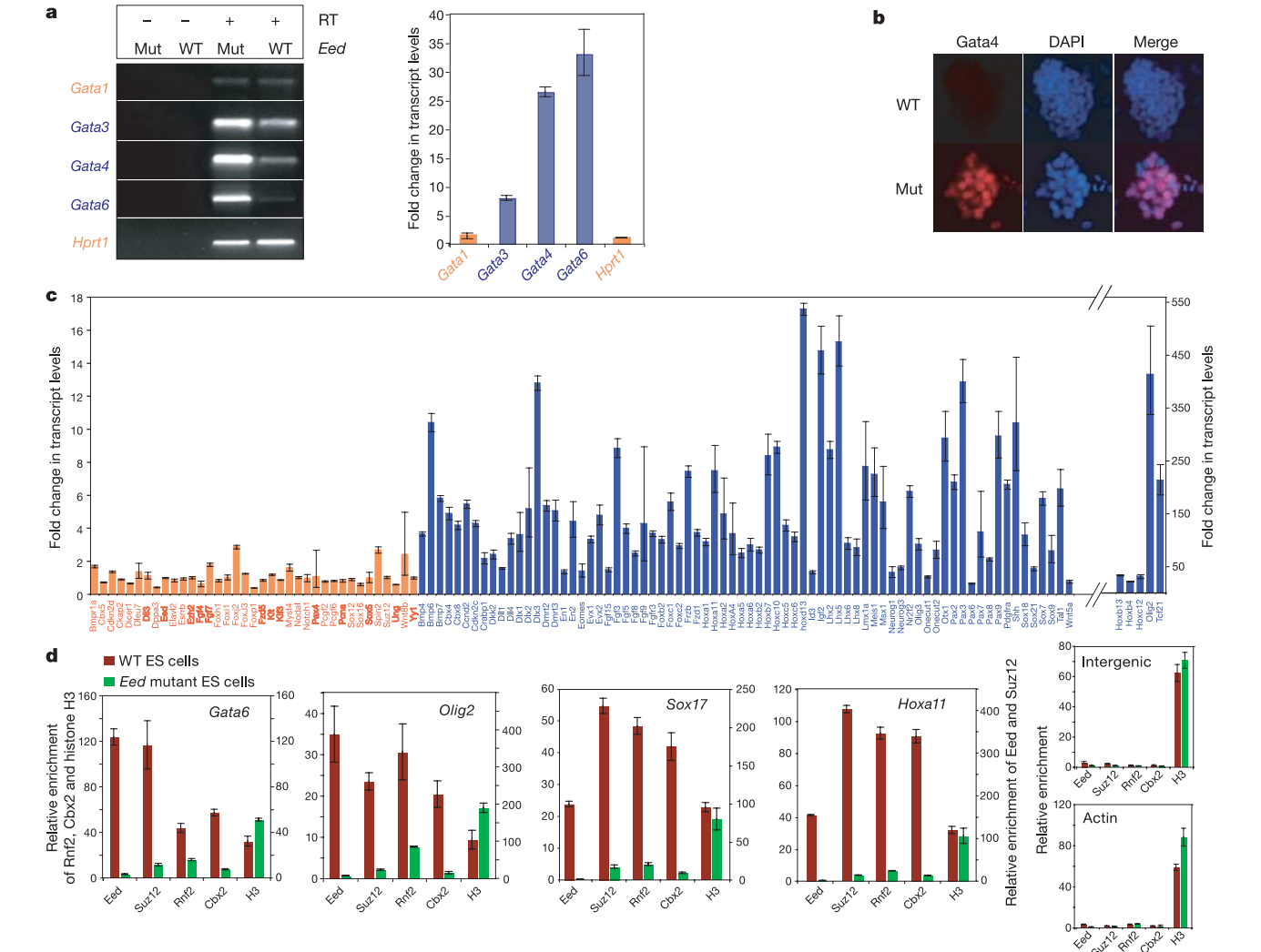


Figure 2 | De-repression of PcG target genes and loss of PRC1 binding in the absence of the PRC2 component *Eed*. **a**, Real-time PCR analysis for *Gata* and *Hprt1* gene transcripts in wild-type (WT) and *Eed* mutant (Mut) ES cells (left panel). In control reactions, reverse transcriptase (RT) was omitted (–). Orange and blue depict unbound and bound genes, respectively, as determined by genome-scale location analysis. Transcript levels were quantified by real-time PCR, normalized to a *Gapdh* control and depicted as a fold change between *Eed* mutant and wild-type ES cells. Error bars are based on the standard deviation derived from triplicate PCR reactions. **b**, Immunostaining for Gata4 in wild-type and *Eed* mutant ES cells. Nuclei

are stained with DAPI. **c**, Quantification of transcript levels in *Eed* mutant ES cells relative to wild-type ES cells, as described in **a**. Genes with the highest relative expression changes are shown at the right of the graph. **d**, Association of PcG components with promoter regions of representative target genes, an intergenic control region, and the highly transcribed actin (*Actb*) gene was determined by ChIP and site-specific real-time PCR in wild-type (brown) and *Eed* mutant (green) ES cells. Histone H3 enrichment levels indicate that an equal amount of input material was used. Error bars represent standard deviations determined from three independent experiments.

Table 1 | Analysis of transcript levels and H3K27 methylation in ES cells and ES-cell-derived NP cells

Group*	Gene	Ratio of transcript levels between NP and ES cells†	Relative H3K27me3 enrichment‡				
			ES cells	U/E	NP cells	U/E	Ratio in NP/ES cells
A	<i>Olig1</i> §	232 (221–245)	186 ± 7	E	2.1 ± 0.2	U	0.01 ± 0.001
	<i>Olig2</i> §	98 (92–105)	518 ± 6	E	1.5 ± 0.4	U	0.003 ± 0.0008
	<i>Olig3</i> §	69 (66–72)	320 ± 26	E	1.4 ± 0.05	U	0.004 ± 0.0004
	<i>Egfr</i> §	109 (105–114)	107 ± 5	E	1.1 ± 0.2	U	0.01 ± 0.0015
	<i>Nes</i> §	248 (237–261)	190 ± 9	E	1.6 ± 1.0	U	0.009 ± 0.005
	<i>Sox9</i> §	719 (690–749)	1021 ± 24	E	1.0 ± 0.2	U	0.001 ± 0.0002
B	<i>Gata6</i> §	0.21 (0.16–0.26)	797 ± 84	E	127 ± 10	E	0.16 ± 0.02
	<i>Hoxa11</i> §	2.02 (1.91–2.14)	442 ± 5	E	58 ± 3	E	0.13 ± 0.01
	<i>Lmx1a</i> §	1.53 (1.21–1.92)	327 ± 79	E	26 ± 0.4	E	0.08 ± 0.02
	<i>Sox17</i> §	1.91 (1.37–2.68)	663 ± 19	E	66 ± 35	E	0.1 ± 0.05
C	<i>Oct4</i>	0.0008 (0.0007–0.0009)	2.1 ± 0.2	U	1.8 ± 0.2	U	0.88 ± 0.12
	<i>Nanog</i>	0.0016 (0.0015–0.016)	7.0 ± 0.5	U	2.6 ± 0.1	U	0.37 ± 0.03
D	<i>Sox2</i>	1.61 (1.57–1.65)	7.8 ± 0.6	U	1.8 ± 0.3	U	0.22 ± 0.04
Other	Intergenic, chr. 8	NA	0.7 ± 0.2	U	2.9 ± 0.8	U	4.25 ± 1.51
	Intergenic, chr. 6	NA	10.6 ± 1.9	U	0.8 ± 0.5	U	0.07 ± 0.05

Chr., chromosome; E, enriched; U, not enriched; NA, not applicable.
* Group A genes are repressed in ES cells and strongly expressed in NP cells. Group B genes are repressed in ES and NP cells. Group C genes are expressed at high levels in ES cells and repressed in NP cells. Group D genes are expressed at high levels in ES and NP cells.
† Determined by real-time PCR with reverse transcription. The error range (in parentheses) is determined based on the standard deviation derived from triplicate PCR reactions.
‡ Determined by ChIP and site-specific real-time PCR. The U/E status of H3K27me3 enrichment at the regions analysed by site-specific PCR is based on the 'not enriched' threshold level determined by intergenic regions. Error ranges indicate s.d. of triplicate PCR reactions.
§ Genes identified in our genome-wide location analysis as PcG and H3K27me3 targets in undifferentiated ES cells.

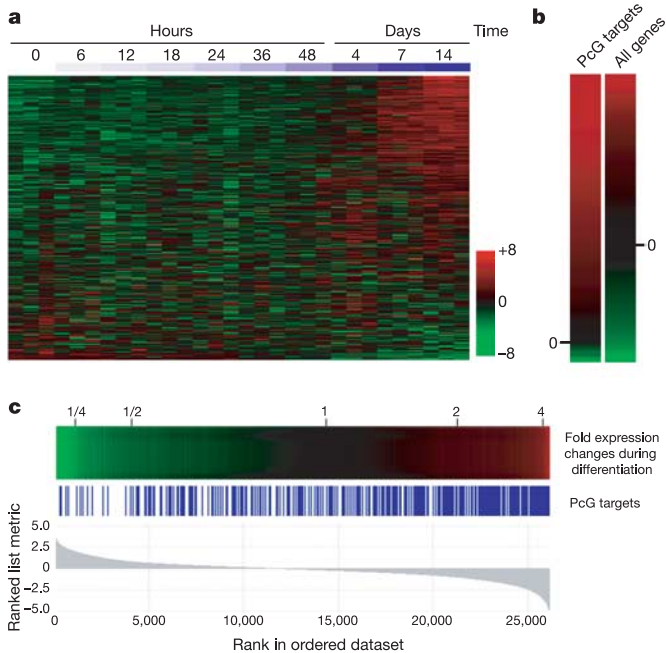


Figure 3 | PcG target genes are preferentially upregulated during ES cell differentiation. **a**, Triplicate expression profiling data sets were analysed at ten time points during the differentiation of ES cells into embryoid bodies. Log₂-transformed probe signal levels are shown relative to the average signal level of each probe set across all samples for PcG target genes. Genes were ordered by the ratio of average signal value in undifferentiated ES cells (0 h) over the average signal value of differentiated ES cells (14 days). Green and red represent lower-than-average and higher-than-average signal levels, respectively. **b**, The ratio of the average signal value for undifferentiated ES cells and ES cells differentiated for 14 days was determined for all probes representing PcG target genes analysed in **a**, and for all probes on the array. Shown is the distribution of log₂-transformed ratios for transcripts that are up- or down-regulated by more than 1.5-fold in both experimental groups. **c**, Triplicate sets of expression profiles were compared between undifferentiated ES cells and cells after 14 days of differentiation. All probe sets in the data set were assigned a score based on a signal-to-noise ratio algorithm (grey bars) and rank-ordered by this score. PcG target genes in the ordered data set are shown as blue lines, and the change in probe signal level is indicated for each probe using the colorimetric scale as in **a**.

METHODS

A detailed description of all materials and methods used can be found in Supplementary Information.
Growth of murine ES cells and derivation of NP cells. E14 (ola/129) ES cells were plated without irradiated murine embryonic fibroblasts (MEFs) and grown under typical ES cell conditions on gelatinized tissue-culture plates. The *Eed* mutant ES cell line (17Rn5-3354SB) obtained from T. Magnuson was grown on irradiated MEFs under standard conditions. V6.5 (C57BL/6-129) ES cells were differentiated along the neural lineage using standard protocols (see Supplementary Information).
Antibodies and chromatin immunoprecipitation assays. Detailed descriptions of antibodies, antibody specificity and ChIP methods used in this study are provided in Supplementary Information. Purified immuno-enriched and input genomic DNA was used for real-time sequence-specific PCR reactions.
Array design and data extraction. The design of the oligonucleotide-based promoter array set and data extraction methods are described in Supplementary Information. The microarrays used in this study were manufactured by Agilent Technologies (<http://www.agilent.com>).
Gene Ontology classification of bound genes. Gene Ontology analysis was performed using BiNGO (<http://www.psb.ugent.be/cbd/papers/BiNGO/>).
Comparing binding and expression data. For expression analysis of PcG target genes during ES cell differentiation, expression data from a time course study comparing V6.5 ES cells with several time points on induction of embryoid body differentiation were retrieved from the NCBI Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE3231.
RNA isolation, real-time PCR and analysis of transcript levels. To determine transcript levels in ES and NP cells, RNA was isolated, reverse-transcribed and subjected to real-time PCR using the SYBR Green PCR master mix and the 7000 ABI Detection System (ABI). Detailed information and all oligonucleotide sequences are provided in Supplementary Information.

Received 30 January; accepted 20 March 2006.
Published online 19 April 2006.

1. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
2. O’Carroll, D. *et al.* The Polycomb-group gene *Ezh2* is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
3. Pasini, D., Bracken, A. P., Jensen, M. R., Lazzarini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071 (2004).
4. Shumacher, A., Faust, C. & Magnuson, T. Positional cloning of a global regulator of anterior–posterior patterning in mice. *Nature* **383**, 250–253 (1996).
5. Voncken, J. W. *et al.* *Rnf2* (*Ring1b*) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl Acad. Sci. USA* **100**, 2468–2473 (2003).

6. Isono, K. *et al.* Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of *Hox* genes. *Mol. Cell. Biol.* **25**, 6694–6706 (2005).
7. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
8. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
9. Czermin, B. *et al.* *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
10. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
11. Muller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
12. Kirmizis, A. *et al.* Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* **18**, 1592–1605 (2004).
13. Hombria, J. C. & Lovegrove, B. Beyond homeosis—HOX function in morphogenesis and organogenesis. *Differentiation* **71**, 461–476 (2003).
14. Schepers, G. E., Teasdale, R. D. & Koopman, P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev. Cell* **3**, 167–170 (2002).
15. Lehmann, O. J., Sowden, J. C., Carlsson, P., Jordan, T. & Bhattacharya, S. S. Fox's in development and disease. *Trends Genet.* **19**, 339–344 (2003).
16. Showell, C., Binder, O. & Conlon, F. L. T-box genes in early embryogenesis. *Dev. Dyn.* **229**, 201–218 (2004).
17. Burch, J. B. Regulation of GATA gene expression during vertebrate development. *Semin. Cell Dev. Biol.* **16**, 71–81 (2005).
18. Montgomery, N. D. *et al.* The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr. Biol.* **15**, 942–947 (2005).
19. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol. Cell* **20**, 845–854 (2005).
20. Gidekel, S. & Bergman, Y. A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element. *J. Biol. Chem.* **277**, 34521–34530 (2002).
21. Hattori, N. *et al.* Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J. Biol. Chem.* **279**, 17063–17069 (2004).
22. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature Cell Biol.* **8**, 188–194 (2006).
23. Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
24. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
25. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
26. Lee, T. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 1–13 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Biology and Research Computing (BaRC), especially B. Yuan, T. Dicesare and T. Volkert, at the Whitehead Institute's Center for Microarray Technology, as well as E. Herbolzheimer, for computational and technical support. We also thank members of the Gifford, Young and Jaenisch laboratories for discussions and critical review of the manuscript. We are grateful to T. Magnuson for the *Eed* mutant (17Rn5-33545B) ES cell lines. L.A.B. was supported in part by an NIH NRSA fellowship. K.P. is a special fellow of the Leukemia and Lymphoma Society. M.W. is supported by a Human Frontiers Science Program Organization Fellowship. M.V. is supported by a grant from the Spanish Ministry of Science and Education. This work was supported in part by grants from the NIHGR1 and NIGMS to R.A.Y. and from the NIH and NCI to R.J.

Author Information All microarray data from this study are available from ArrayExpress at the EBI (<http://www.ebi.ac.uk/arrayexpress>) under accession code E-WMIT-12. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J. (jaenisch@wi.mit.edu).

LETTERS

A TAS1R receptor-based explanation of sweet 'water-taste'

Veronica Galindo-Cuspinera¹, Marcel Winnig², Bernd Bufe², Wolfgang Meyerhof² & Paul A. S. Breslin¹

'Water-tastes' are gustatory after-impressions elicited by water following the removal of a chemical solution from the mouth, akin to colour after-images appearing on 'white' paper after fixation on coloured images. Unlike colour after-images, gustatory after-effects are poorly understood¹. One theory posits that 'water-tastes' are adaptation phenomena, in which adaptation to one taste solution causes the water presented subsequently to act as a taste stimulus^{2,3}. An alternative hypothesis is that removal of the stimulus upon rinsing generates a receptor-based, positive, off-response in taste-receptor cells, ultimately inducing a gustatory perception⁴. Here we show that a sweet 'water-taste' is elicited when sweet-taste inhibitors are rinsed away. Responses of cultured cells expressing the human sweetener receptor directly parallel the psychophysical responses—water rinses remove the inhibitor from the heteromeric sweetener receptor TAS1R2–TAS1R3, which activates cells and results in the perception of strong sweetness from pure water. This 'rebound' activity occurs when equilibrium forces on the two-state allosteric sweet receptors result in their coordinated shift to the activated state upon being released from inhibition by rinsing^{5–7}.

A readily available sweet 'water-taste' stimulus is sodium saccharin (Na-saccharin). Na-saccharin is a non-caloric sweetener commonly used at low concentrations in foods and beverages. At high concentrations, however, Na-saccharin elicits little sweetness and tastes mostly bitter after consecutive exposures⁸ (Fig. 1a). When high concentrations are rinsed from the mouth, the perception switches from a primarily bitter taste to an intense sweet taste (Fig. 1b).

The concentration–intensity curve for Na-saccharin showed that sweetness perception initially increased with concentration, reaching a plateau at ~4 mM, whereas bitterness increased slowly (Supplementary Fig. S1a). Upon a second tasting of the same concentration, sweetness was markedly reduced at concentrations above 6.25 mM, which suggests that this compound acts as a sweet agonist at low concentrations and as an antagonist at high concentrations (Fig. 1a).

From the 'water-taste' function, we see that sweet 'water-taste' first appeared when 6.25 mM Na-saccharin was rinsed from the mouth (Fig. 1b). It should be noted that sweet 'water-taste' does not refer to a lingering taste from rinsing a sweetener from the mouth, as water rinses after most sweeteners taste 'barely' sweet (Fig. 1d, right panel). For each concentrated Na-saccharin solution, the intensity of sweet 'water-taste' that followed it was proportional to the degree of apparent sweetness inhibition that occurred during sampling (Fig. 1a, b). On the basis of this relationship, we postulated that other sweet 'water-taste' compounds might also be sweet-taste inhibitors.

To determine whether compounds that elicit sweet 'water-taste' also inhibit sweetness, we tested lactisole, MgSO₄ and high concentrations of Na-saccharin and acesulfame-K, all of which elicit sweet

'water-taste' (Fig. 1c, d, middle panels). Subjects tasted several sweeteners alone and in binary mixtures with these compounds. All four sweet 'water-taste' stimuli inhibited the moderate sweetness intensity of 300 mM sucrose, 4 mM Na-saccharin, 17.5 mM Na-cyclamate and 3 mM aspartame (Fig. 2a–d). (Concentrations of sweetener were initially matched in terms of sweetness intensity⁹; Supplementary Fig. S2a.) Even when tasted alone, high concentrations of acesulfame-K and Na-saccharin self-inhibit their sweet taste, suggesting a similar mode of action for these two compounds (Fig. 1a, c, middle and right panels, and Fig. 2a, d). In contrast, Na-cyclamate, a sweetener with no sweet 'water-taste', did not inhibit sweetness when mixed with other sweeteners (Fig. 2f). Together, these findings provide strong

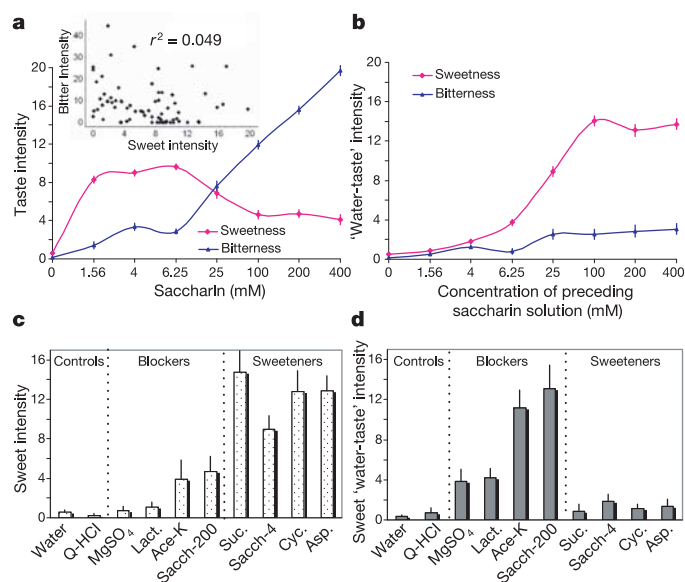


Figure 1 | Concentration-intensity curves and sweet 'water-taste'.

a, Concentration–intensity relationship for Na-saccharin, measured during the second tasting of a given concentration ($n = 14$ subjects). Inset shows the bitter–sweet correlation for Na-saccharin, with pooled data from 1.5–400 mM. Intensity was rated on the gLMS scale (see Methods). **b**, 'Water-taste' concentration–intensity curve for Na-saccharin. The taste intensity of deionized water was measured after two consecutive exposures to the same Na-saccharin solution (10-s delay between exposures, $n = 14$ subjects). **c**, Sweetness intensity measured during the second exposure to individual compounds. **d**, Sweet 'water-taste' from each compound, tasted with 10-s delay. Compounds tested were 0.15 mM quinine-HCl (Q-HCl), 500 mM MgSO₄, 1 mM lactisole (Lact.), 200 mM acesulfame-K (Ace-K), 200 mM Na-saccharin (Sacch-200), 300 mM sucrose (Suc.), 4 mM Na-saccharin (Sacch-4), 17.5 mM Na-cyclamate (Cyc.) and 3 mM aspartame (Asp.). Error bars indicate s.e.m.

¹Monell Chemical Senses Center, 3500 Market Street, Philadelphia, Pennsylvania 19104, USA. ²German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany.

evidence that sweet 'water-taste' is associated with compounds that inhibit sweet taste.

However, the cognitive suppression of sweetness by bitter taste could provide an alternative explanation for sweet-taste inhibition^{8,10}. Na-saccharin, acesulfame-K and MgSO₄ have a bitter side-taste (Fig. 1a and Supplementary Fig. 2b), but the perceived sweetness of Na-saccharin does not decrease at the same rate that bitterness perception increases with concentration (Fig. 1a and Supplementary Fig. 1a). Furthermore, the degree of Na-saccharin bitterness at a given concentration does not correlate with the degree of sweetness ($r^2 = 0.049$, inset in Fig. 1a; Supplementary Fig. 1b). In addition, quinine-HCl, used as a control matched for bitterness intensity (Supplementary Fig. S2b), does not significantly suppress sweetness in mixtures (Fig. 2e).

Adaptation of the sweet taste could also account for why sweetness perceptions of Na-saccharin and acesulfame-K solutions were reduced upon second tasting¹¹. To test this hypothesis, we repeatedly presented subjects with a solution of 200 mM Na-saccharin mixed with 580 mM sucrose (Fig. 3). Subjects were first provided with a 580 mM sucrose solution, which they described as 'moderate' to 'strong' (~18 on the general labelled magnitude scale). The second solution they tasted was the mixture, which was rated as 'moderately' sweet. With subsequent presentations of the same mixture, sweetness perception was reduced to a weak level, where it remained until subjects tasted sucrose without Na-saccharin at the end of the series. The immediate and strong sweet taste perceived from sucrose alone shows that no sucrose taste-adaptation had occurred (Fig. 3, compare trials 1 and 10). The failure of sucrose sweetness to adapt (when

adaptation would ordinarily occur in the absence of inhibition¹¹) suggests that sucrose was not activating the sweet-taste system when mixed with high Na-saccharin concentrations because Na-saccharin inhibits activation of the sweet-taste transduction system.

In order to determine whether Na-saccharin directly inhibits the human sweetener receptor, we used an *in vitro* HEK293T cell system to heterologously express the human sweet receptor heteromer (h)TAS1R2–(h)TAS1R3 (refs 12, 13). We monitored the response of receptor-expressing cells when exposed to different concentrations of Na-saccharin by calcium imaging using a fluorescence imaging plate reader (FLIPR). Application of Na-saccharin or acesulfame-K elicited a dose-dependent response that saturated at 3 mM, and concentrations above 10 mM inhibited the response to both sweeteners (Fig. 4a). Notably, this inhibitory effect started at the same concentration at which it occurred perceptually in humans (Figs 1a and 4a).

The inhibition seen with these sulfonyl-amide sweeteners is particular to the stimulus–receptor combinations used here. The stimulus specificity is illustrated by comparing sweetener–inhibitor functions with the dose–response curve for the sweetener Na-cyclamate, which does not decline at high concentrations in cells expressing the sweet-taste receptor (Supplementary Fig. 3d). Receptor specificity is illustrated by Na-saccharin or acesulfame-K activation of cells expressing their cognate bitter-taste receptor hTAS2R44 (ref. 14), which increases monotonically with concentration (Fig. 4b). Unlike the human sweet receptor, the response of cells expressing the rat sweetener receptor rTas1r2–rTas1r3 did not decline at Na-saccharin concentrations up to 60 mM (Supplementary Fig. S3c, e). Thus, inhibition by Na-saccharin is specific to the human hTAS1R2–hTAS1R3 receptor. Furthermore, a chimaeric receptor consisting of the transmembrane region of human TAS1R3 attached to the amino-terminal region of rat Tas1r3 co-expressed with rat Tas1r2 was inhibited by high Na-saccharin concentrations, but the reverse chimaera was not (Supplementary Fig. S4b–e). This suggests that the allosteric binding site for inhibition by Na-saccharin, like that for the sweetness inhibitor and 'water-taste' compound lactisole^{15–17}, is located in the transmembrane region of hTAS1R3.

We also wanted to determine whether high Na-saccharin concentrations would inhibit other sweeteners *in vitro* when mixed together. Similar to the results of our human psychophysical experiments,

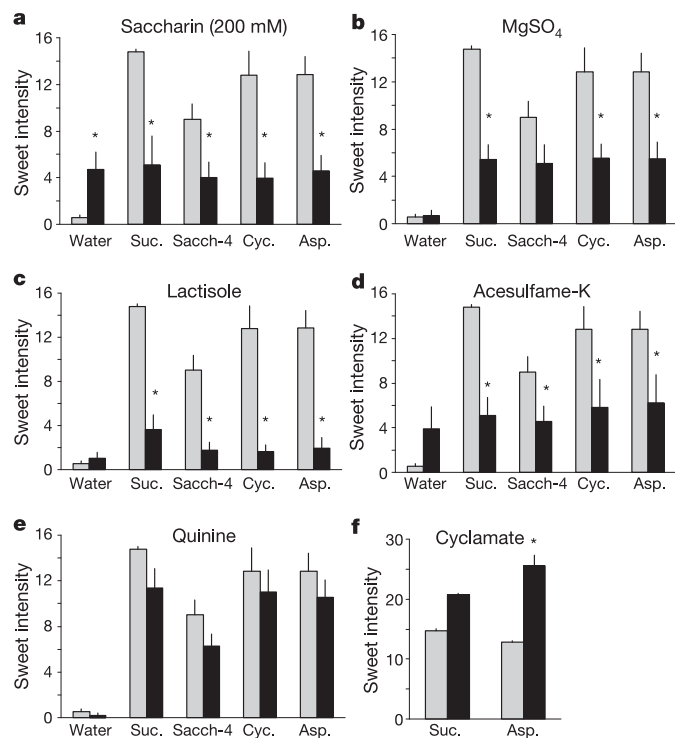


Figure 2 | Sweet-blocking effects of sweet 'water-taste' stimuli. a–f, Sweet-blocking effects of 200 mM Na-saccharin (a), 500 mM MgSO₄ (b), 1 mM lactisole (c), 200 mM acesulfame-K (d), 0.15 mM quinine-HCl (e) and 100 mM Na-cyclamate (f). Grey bars refer to the mean intensity of the pure sweeteners listed on the x axis ($n = 14$); black bars reflect the mean sweetness for mixtures of sweetener and putative blocker. The agonists tested were 300 mM sucrose (Suc.), 4 mM Na-saccharin (Sacch-4), 17.5 mM Na-cyclamate (Cyc.) and 3 mM aspartame (Asp.). Quinine-HCl was used as a bitter control and Na-cyclamate (100 mM) was a control for high sweetener concentration. Error bars indicate s.e.m. Asterisks indicate statistical significance at $\alpha = 0.05$ using repeated measures ANOVA.

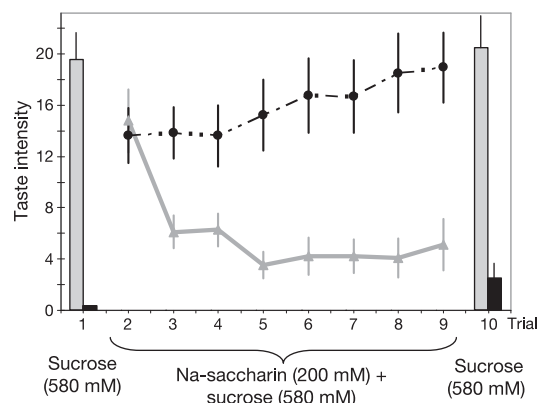


Figure 3 | Blocking versus adaptation of sweet-taste by saccharin. Mean sweetness and bitterness of 20% sucrose (580 mM) and an Na-saccharin:sucrose (200 mM:580 mM) mixture. Fourteen subjects tasted the solutions as shown, at 10-s intervals (90 s total), and rated sweetness and bitterness on a gLMS scale without rinsing between tastings. Grey bars indicate sweetness intensity of sucrose, black bars indicate sucrose bitterness. Triangles indicate the sweetness of the Na-saccharin:sucrose mixture; circles indicate the bitterness of the mixture. Error bars indicate s.e.m.

Na-saccharin inhibited the responses of hTAS1R2–hTAS1R3-expressing cells to all sweeteners tested (Fig. 4c).

To determine whether the observed Na-saccharin inhibition of the hTAS1R2–hTAS1R3 heteromer could explain the ‘water-taste’, we replicated the inhibitor rinse-off *in vitro*. We used single-cell calcium imaging for these experiments, as the FLIPR bath solutions cannot be easily exchanged. We challenged cells expressing hTAS1R2–hTAS1R3 with 50 mM Na-saccharin and subsequently rinsed the cells (Fig. 4d). Although there was almost no response to application of 50 mM Na-saccharin, a robust response was triggered upon rinsing—this response directly paralleled the sweet ‘water-taste’ detected perceptually (Fig. 1b, d). In contrast, application of 3 mM Na-saccharin readily activated hTAS1R2–hTAS1R3, but rinsing had no effect, which is consistent with the perception of low concentrations of Na-saccharin (Figs 1a, b and 4d). Similar single-cell responses to water rinses were also obtained with acesulfame-K (Fig. 4e and Supplementary Fig. S5b, c). The specificity of the rinse for hTAS1R2–hTAS1R3-expressing cells is further demonstrated by the observation that the bitter-taste receptor hTAS2R44 responded robustly to 50 mM Na-saccharin, whereas the rinse did not elicit a

response in these or in mock-transfected control cells (Supplementary Fig. S5a, d).

To demonstrate that sweet ‘water-taste’ is specific to the rinsing out of sweet receptor inhibitors rather than rinsing out of ‘sweeteners’ *per se*, we examined the effects of the sweet ‘water-taste’ compound lactisole. Lactisole does not directly elicit sweet taste perceptually¹⁸ and does not activate the hTAS1R2–hTAS1R3 receptor at any of the concentrations we tested (Supplementary Fig. S6c). In contrast, lactisole decreased basal intracellular calcium levels in single-cell imaging experiments (Fig. 4f, g and Supplementary Fig. S6a). However, when rinsed away, lactisole triggered a response in hTAS1R2–hTAS1R3-expressing cells but not in control cells, which directly parallels the sweet ‘water-taste’ elicited by rinsing off lactisole in human taste experiments (Figs 1d, 4f and Supplementary Fig. S6a, b). Receptor activation upon rinsing away of lactisole further illustrates that sweet ‘water-taste’ arises from the removal of sweet-receptor inhibitors.

The effects of Na-saccharin and acesulfame-K on the hTAS1R2–hTAS1R3 heteromer can be best explained by an allosteric receptor model^{19–21}. At low concentrations, Na-saccharin and acesulfame-K bind preferentially to the high-affinity activating site on the receptor. At high concentrations, they also bind to a lower affinity allosteric site that shifts the receptor equilibrium to an inactive receptor conformation^{22,23}, thereby inhibiting sweet taste. Sweet ‘water-taste’ can be explained by the preferential removal of these compounds from the low-affinity inhibitory site upon rinsing.

Although lactisole causes no directly observable activation of the sweet-receptor (Supplementary Fig. S6c), we believe that it stimulates sweet ‘water-taste’ by a mechanism very similar to Na-saccharin and acesulfame-K. The hTAS1R2–hTAS1R3 receptor is constitutively active, increasing basal intracellular calcium levels when expressed in HEK293 cells (Fig. 4g). Exposing hTAS1R2–hTAS1R3-expressing cells to lactisole reduces calcium concentrations to below basal levels (compare Fig. 4g and Fig. 4d, e for related effects of Na-saccharin and acesulfame-K), supporting our hypothesis that lactisole is an inverse agonist that moves receptor equilibrium from a constitutively active to an inhibited conformational state^{5,15,16,24–28}. Removing lactisole induces a coordinated shift of the receptors from the high-energy, inhibited state to the constitutively active state, triggering the observed cellular response (Fig. 4f and Supplementary Fig. 6a)—which in turn results in the perception of sweet ‘water-taste’.

Our combined results suggest that sweet ‘water-taste’ can be used to predict potential sweet-taste blockers. Our data also show that the commercial sweeteners Na-saccharin and acesulfame-K are, paradoxically, sweet-taste inhibitors at high concentrations, which causally links sweet ‘water-taste’ to sweetness inhibition in a two-state allosteric model.

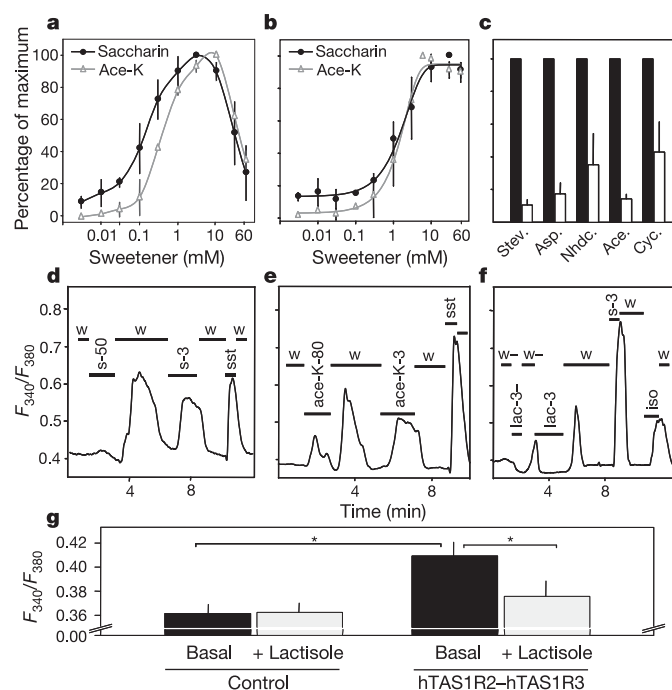


Figure 4 | Functional properties of hTAS1R2–hTAS1R3 and hTAS2R44.

a, b, FLIPR calcium responses averaged over three independent experiments ($n = 3$) to Na-saccharin (filled circles) and acesulfame-K (open triangles) of cells expressing hTAS1R2–hTAS1R3 (**a**) or hTAS2R44 (**b**). **c**, Responses of cells expressing hTAS1R2–hTAS1R3 to 1 mM stevioside (Stev.), 5 mM aspartame (Asp.), 1 mM neohesperidine dihydrochalcone (Nhdc), 5 mM acesulfame-K (Ace.) and 5 mM Na-cyclamate (Cyc.) in the presence (white) or absence (black) of 60 mM Na-saccharin, normalized to the response elicited by the pure sweetener. **d–f**, Typical single-cell imaging traces of cells expressing hTAS1R2–hTAS1R3, to examine effects of rinsing off Na-saccharin (**d**, 40 responders out of 108 cells, $n = 3$), acesulfame-K (**e**, 33 responders of 98 cells, $n = 1$) and lactisole (**f**, 32 responders of 61 cells, $n = 3$). Treatment conditions were wash (w), 50 mM Na-saccharin (s-50), 3 mM Na-saccharin (s-3), 80 mM acesulfame-K (ace-K-80), 3 mM acesulfame-K (ace-K-3), 3 mM lactisole (lac-3), 10 μ M isoproterenol (iso) and 0.1 μ M SST14 (sst). Minus sign indicates the absence of 5 mM glucose in the medium. Horizontal bars indicate the duration of the treatment. **g**, Basal calcium levels in cells constitutively expressing $G_{\alpha 15}G_{\beta 3}$ and hTAS1R2, with or without hTAS1R3 expression, in the presence and absence of 3 mM lactisole (12 cells, $n = 2$). Asterisks indicate $P < 0.001$ (t -test). Error bars denote s.d.

METHODS

Human psychophysics. Fourteen subjects (ten females) aged 24–51 were presented with trials consisting of three cups, with the first two cups containing the same stimulus, and the third cup containing deionized water. The inter-stimulus delay was 10 s, and the time per trial ~ 1 min. Test stimuli were swished for approximately 5 s in the mouth and then expectorated. Subjects were then asked to rate the intensity of the five qualities of taste (sweet, sour, salty and umami) for each of the three samples on a general labelled magnitude scale (gLMS), and were asked not to rinse between solutions. The minimum time between trials was 1 h. The trial structure was designed to quantify any changes in taste-quality intensities that may occur with repeated tasting, and to measure the ‘water-taste’ of individual stimuli.

Subjects were initially trained to use the gLMS scale according to standard published procedures²⁹, with the top of the scale described as the ‘strongest imaginable’ sensation of any kind³⁰. Each subject participated in 72 sessions over a period of 24 days, and was asked not to eat, drink or chew gum 1 h before each session. Stimuli in the ‘sweetener’ category were sucrose (300 mM), *N*-L- α -aspartyl-L-phenylalanine (aspartame, 3 mM), Na-cyclamate (17.5 mM) and Na-saccharin (saccharin-A, 4 mM). Acesulfame-K (200 mM), Na-($\pm$)-2-(*P*-methoxy-phenoxy)-propionic-acid (lactisole, 1.53 mM), magnesium sulphate 99% ($MgSO_4$, 500 mM) and Na-saccharin (saccharin-B, 200 mM) were used as

'blockers'. HCl-quinine (0.151 mM) was used as a control for bitterness intensity and Na-cyclamate (100 mM) as a control for high sweetener concentration. Analytical grade compounds were obtained from Sigma. Solutions were prepared in sterile deionized water (Millipore-filtered) and presented at 21 °C in 40-ml plastic cups. Lactisole solutions were adjusted to pH 7.0 by adding 0.1 N NaOH.

Na-saccharin psychophysical curves. Six concentrations of Na-saccharin were tested (1.6 mM, 6.2 mM, 25 mM, 100 mM, 200 mM and 400 mM) in random order and in duplicate, following the described protocol.

'Water-taste' and sweet-blocking effects. All stimuli were tested in duplicate and in random order, following the described protocol. Stimuli were presented individually and in all possible sweetener/blocker combinations.

Blocking versus adaptation effects. A series of Na-saccharin:sucrose mixtures were presented to subjects at 10-s intervals. Trials consisted of ten cups: the first containing a sucrose solution (580 mM), followed by eight cups containing a mixture of Na-saccharin:sucrose (200 mM:580 mM) and finishing with a cup of 580 mM sucrose. Subjects were asked to rate the sweetness and bitterness intensities of each solution on a gLMS scale, and were not allowed to rinse between samples. The whole experiment was replicated in the same subjects.

Statistical analysis. Repeated measures analysis of variance (ANOVA) followed by Tukey's pairwise post-hoc comparisons were performed on the standardized data using Statistica v.6.1. A Bonferroni correction was used only when testing individual compounds for sweet 'water-taste', using $\alpha = 0.05$. Simple linear regression using the general linear model module (GLM) was used to test for correlations between Na-saccharin sweetness and bitterness.

Ca²⁺-FLIPR assay. cDNAs for the sweet and the bitter taste receptors were transiently transfected into HEK293T cells expressing G_{α16gust44} (ref. 16). To avoid osmotic effects, we substituted the sodium in the Cl solution (130 mM NaCl, 5 mM KCl, 10 mM HEPES, 2 mM CaCl₂, 5 mM glucose, pH 7.4) with Na-saccharin and Na-cyclamate in concentrations above 10 mM proportional to the concentrations in which they were used. Amplitudes of calcium signals were corrected for the response of mock-transfected cells and normalized to the fluorescence of cells before the stimulus using $\Delta F/F = (F - F_0)/F_0$. Dose-response curves and EC₅₀ values were calculated by nonlinear regression¹⁶.

Single-cell Ca²⁺-imaging. Single-cell Ca²⁺-imaging was performed using a Till Photonics monochromatic-based imaging system in combination with an Olympus IX50 inverted microscope equipped with a UApo/340 40 × 1.35 oil immersion lens. To investigate the effects of Na-saccharin rinse-off, cDNAs for human TAS1R2–TAS1R3 or TAS2R44 were transiently transfected into HEK293T/G_{α16gust44} cells¹⁶. To investigate the effect of rinsing lactisole off the receptor, we used a HEK293 FlpIn T-REx cell line (Invitrogen) constitutively expressing hTAS1R2 and the G-protein chimera G_{α15}G₁₃ (containing the last five amino acids of human G₁₃), with TAS1R3 expression under the control of a tetracycline-responsive element. Twenty-four hours before the experiment, hTAS1R3 expression was induced using 0.5 μg ml⁻¹ tetracycline. Cells were loaded with the Ca²⁺-sensitive fluorescent dye Fura-2-AM (4 μM, Molecular Probes) for 1 h at 37 °C, and then washed with Cl solution. For each experiment, a glass coverslip with Fura-2-loaded cells was placed in a microperfusion chamber. The cells were superfused with test solutions for 80–100 s (5 ml per application) and washed out with 5 ml Cl solution at the end of each application. All test solutions were manually applied to the chamber using a micropipette. At the end of each experiment, somatostatin-14 (SST14, 0.1 μM) or isoproterenol (10 μM) was applied to stimulate endogenous somatostatin or β-adrenergic receptors, respectively, proving that the G-protein dependent signal transduction cascade was functional. The fluorescence ratio (F_{340}/F_{380}) in single cells was monitored at 5-s intervals at an emission wavelength of 515 nm, after excitation (3–15 ms with 340 nm and 380 nm, using an intensified, cooled CCD camera). Signals from all responding cells or all cells (negative control) were averaged and displayed as a function of time.

Received 19 December 2005; accepted 31 March 2006.

Published online 23 April 2006.

1. Bartoshuk, L. M. Water taste in man. *Percept. Psychophys.* **3**, 69–72 (1968).
2. McBurney, D. H. & Bartoshuk, L. M. in *Olfaction and taste IV* (ed. Schneider, D.) 329–335 (Wissenschaftliche Verlagsgesellschaft MBH, Stuttgart, 1972).
3. McBurney, D. H. & Bartoshuk, L. M. Interactions between stimuli with different taste qualities. *Physiol. Behav.* **10**, 1101–1106 (1973).
4. DuBois, G. E. Unraveling the biochemistry of sweet and umami tastes. *Proc. Natl Acad. Sci. USA* **101**, 13972–13973 (2004).
5. Carroll, F. Y. et al. BAY36-7620: a potent non-competitive mGlu1 receptor antagonist with inverse agonist activity. *Mol. Pharmacol.* **59**, 965–973 (2001).
6. Kasai, H. Tonic inhibition and rebound facilitation of a neuronal calcium channel by a GTP-binding protein. *Proc. Natl Acad. Sci. USA* **88**, 8855–8859 (1991).

7. Harvey, R. D. & Belevych, A. E. Muscarinic regulation of cardiac ion channels. *Br. J. Pharmacol.* **139**, 1074–1084 (2003).
8. Horne, J., Lawless, H. T., Speirs, W. & Sposato, D. Bitter taste of saccharin and acesulfame-K. *Chem. Senses* **27**, 31–38 (2002).
9. Schiffman, S. S. & Gatlin, C. A. Sweeteners: State of knowledge review. *Neurosci. Biobehav. Rev.* **17**, 313–345 (1993).
10. Schiffman, S. S., Booth, B. J., Losee, M. L., Pecore, S. D. & Warwick, Z. S. Bitterness of sweeteners as a function of concentration. *Brain Res. Bull.* **36**, 505–513 (1995).
11. Gent, J. F. & McBurney, D. H. Time course of gustatory adaptation. *Percept. Psychophys.* **23**, 171–175 (1978).
12. Li, X. et al. Human receptors for sweet and umami taste. *Proc. Natl Acad. Sci. USA* **99**, 4692–4696 (2002).
13. Nelson, G. et al. Mammalian sweet taste receptors. *Cell* **106**, 381–390 (2001).
14. Kuhn, C. et al. Bitter taste receptors for saccharin and acesulfame K. *J. Neurosci.* **24**, 10260–10265 (2004).
15. Jiang, P. et al. Lactisole interacts with the transmembrane domains of human T1R3 to inhibit sweet taste. *J. Biol. Chem.* **280**, 15238–15246 (2005).
16. Winnig, M., Bufe, B. & Meyerhof, W. Valine 738 and lysine 735 in the fifth transmembrane domain of rTas1r3 mediate insensitivity towards lactisole of the rat sweet taste receptor. *BMC Neurosci.* **6**, 1–7 (2005).
17. Xu, H. et al. Different functional roles of T1R subunits in the heteromeric taste receptors. *Proc. Natl Acad. Sci. USA* **101**, 14258–14263 (2004).
18. Schiffman, S. S., Booth, B. J., Sattely-Miller, E. A., Graham, B. G. & Gibes, K. M. Selective inhibition of sweetness by the sodium salt of ± 2 -(4-methoxyphenoxy)propanoic acid. *Chem. Senses* **24**, 439–447 (1999).
19. Dalziel, K. A kinetic interpretation of the allosteric model of Monod, Wyman and Changeaux. *FEBS Lett.* **1**, 346–348 (1968).
20. Kunishima, N. et al. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–977 (2000).
21. Samama, P., Pei, G., Costa, T., Cotecchia, S. & Lefkowitz, R. J. Negative antagonists promote an inactive conformation of the beta 2-adrenergic receptor. *Mol. Pharmacol.* **45**, 390–394 (1994).
22. Cooperman, B. S. in *Encyclopedia of Biological Chemistry* (eds Lennarz, W. J. & Lane, M. D.) 68–73 (Elsevier, New York, 2004).
23. Kashlan, O. B., Scott, C. P., Lear, J. D. & Cooperman, B. S. A comprehensive model for the allosteric regulation of mammalian ribonucleotide reductase. Functional consequences of ATP- and dATP-induced oligomerization of the large subunit. *Biochemistry* **41**, 462–474 (2002).
24. Christopoulos, A. & Kenakin, T. G protein-coupled receptor allostery and complexing. *Pharmacol. Rev.* **54**, 323–374 (2002).
25. Costa, T. & Cotecchia, S. Historical review: Negative efficacy and the constitutive activity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **26**, 618–624 (2005).
26. Jiang, P. et al. Identification of the cyclamate interaction site within the transmembrane domain of the human sweet taste receptor subunit T1R3. *J. Biol. Chem.* **280**, 34296–34305 (2005).
27. Kenakin, T. G-protein coupled receptors as allosteric machines. *Receptors Channels* **10**, 51–60 (2004).
28. Pin, J. P., Galvez, T. & Prezeau, L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol. Ther.* **98**, 325–354 (2003).
29. Green, B. G., Dalton, P., Cowart, B. J., Shaffer, G. S. & Rankin, K. Evaluating the 'labelled magnitude scale' for measuring sensations of taste and smell. *Chem. Senses* **21**, 323–334 (1996).
30. Bartoshuk, L. M. Comparing sensory experience across individuals: recent psychophysical advances illuminate genetic variation in taste perception. *Chem. Senses* **25**, 447–460 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank E. Chudoba, E. Schöley-Pohl and S. Demgensky for technical assistance. We thank G. Beauchamp, B. Cooperman and D. Krautwurst for comments and discussion. We are grateful to H. Schmale for the rat *Tas1r3* cDNA and J. Slack for the HEK293T G_{α16gust44}-expressing cell line. This research was supported in part by NIH and US-Israel BARD grants to P.A.S.B. and the German Science Foundation (W.M.).

Author Contributions V.G.-C. and P.A.S.B. discovered the phenomenon of sweet 'water-taste' with Na-saccharin and other sweetness inhibitors, conceived and designed the perceptual experiments and collected the behavioural data. M.W. conducted the molecular, histological and imaging experiments. B.B. participated in the functional studies and coordination of the molecular experiments. All authors contributed to the design of parallel *in vivo/ex vivo* control experiments and to writing and editing the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.A.S.B. (breslin@monell.org).

LETTERS

Platensimycin is a selective FabF inhibitor with potent antibiotic properties

Jun Wang^{1*}, Stephen M. Soisson^{1*}, Katherine Young¹, Wesley Shoop^{1†}, Srinivas Kodali¹, Andrew Galgoci¹, Ronald Painter¹, Gopalakrishnan Parthasarathy¹, Yui S. Tang¹, Richard Cummings¹, Sookhee Ha¹, Karen Dorso¹, Mary Motyl¹, Hiranthi Jayasuriya¹, John Ondeyka¹, Kithsiri Herath¹, Chaowei Zhang¹, Lorraine Hernandez¹, John Allocco¹, Ángela Basilio¹, José R. Tormo¹, Olga Genilloud¹, Francisca Vicente¹, Fernando Pelaez¹, Lawrence Colwell¹, Sang Ho Lee¹, Bruce Michael¹, Thomas Felcetto¹, Charles Gill¹, Lynn L. Silver^{1‡}, Jeffery D. Hermes¹, Ken Bartizal¹, John Barrett^{1‡}, Dennis Schmatz¹, Joseph W. Becker¹, Doris Cully¹ & Sheo B. Singh¹

Bacterial infection remains a serious threat to human lives because of emerging resistance to existing antibiotics. Although the scientific community has avidly pursued the discovery of new antibiotics that interact with new targets, these efforts have met with limited success since the early 1960s^{1,2}. Here we report the discovery of platensimycin, a previously unknown class of antibiotics produced by *Streptomyces platensis*. Platensimycin demonstrates strong, broad-spectrum Gram-positive antibacterial activity by selectively inhibiting cellular lipid biosynthesis. We show that this anti-bacterial effect is exerted through the selective targeting of β -ketoacyl-(acyl-carrier-protein (ACP)) synthase I/II (FabF/B) in the synthetic pathway of fatty acids. Direct binding assays show that platensimycin interacts specifically with the acyl-enzyme intermediate of the target protein, and X-ray crystallographic studies reveal that a specific conformational change that occurs on acylation must take place before the inhibitor can bind. Treatment with platensimycin eradicates *Staphylococcus aureus* infection in mice. Because of its unique mode of action, platensimycin shows no cross-resistance to other key antibiotic-resistant strains tested, including methicillin-resistant *S. aureus*, vancomycin-intermediate *S. aureus* and vancomycin-resistant enterococci. Platensimycin is the most potent inhibitor reported for the FabF/B condensing enzymes, and is the only inhibitor of these targets that shows broad-spectrum activity, *in vivo* efficacy and no observed toxicity.

Bacterial type II fatty-acid synthesis (FASII) is an attractive target for antibacterial drug discovery^{3–7}. The initiation condensing enzyme, FabH, and elongation condensing enzymes, FabF/B, are essential components of fatty-acid biosynthesis^{8–11} and are highly conserved among key pathogens. Although no drugs targeting condensing enzymes are used clinically, cerulenin¹² and thiolactomycin¹³ selectively inhibit the condensation enzymes FabF/B and FabH.

Systematic screening of 250,000 natural product extracts (83,000 strains in three growth conditions), with the use of a combination of target-based whole-cell and biochemical assays, led to the identification of a potent and selective small molecule from a strain of *Streptomyces platensis* recovered from a soil sample collected in South Africa. This molecule, platensimycin (C₂₄H₂₇NO₇, relative molecular mass 441.47), comprises two distinct structural elements connected by an amide bond (Fig. 1a).

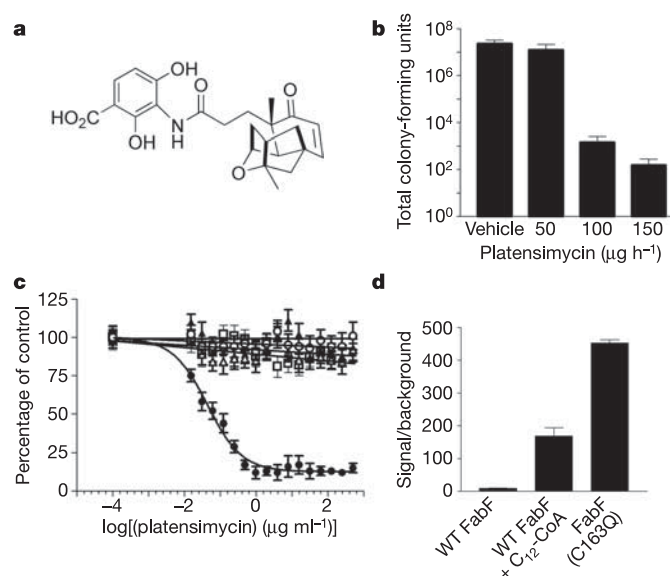


Figure 1 | Characterization of platensimycin. **a**, Structure of platensimycin. **b**, *In vivo* studies on platensimycin. Dosing at 50 µg h⁻¹ showed small decrease in viable *S. aureus* cells from the infected kidney. However, a 10⁴–10⁵ fold decrease (4 and 5 log reduction) were achieved with 100 and 150 µg h⁻¹, respectively. Dosing at 150 µg h⁻¹ showed 40% of the kidneys with no viable *S. aureus*, whereas dosing at 100 µg h⁻¹ showed 20% of the kidneys without detectable viable *S. aureus*. Error bars indicate s.d. observed with five infected mice. The results were confirmed by a repeat experiment. **c**, Whole-cell labelling assay¹⁶ with platensimycin. The assay was performed with a serial dilution of platensimycin, starting at 500 µg ml⁻¹. Platensimycin showed no significant inhibition against syntheses of DNA (open circles), cell wall (filled triangles), protein (open squares) and RNA (open triangles) but greatly inhibited phospholipid synthesis (filled circles), providing an IC₅₀ value of 0.1 µg ml⁻¹. Error bars indicate s.d. for three individual experiments. **d**, Direct binding assay results of [³H]dihydroplatensimycin and *E. coli* FabF (ecFabF) in the presence and absence of n-dodecanoyl coenzyme A (lauroyl-CoA; C₁₂-CoA) and the C163Q mutant protein. Error bars indicate s.d. observed with six replicate wells. Experimental details are given in Supplementary Information.

¹Merck Research Laboratories, Rahway, New Jersey 07065, USA. [†]Present addresses: DuPont Stine-Haskell Research Center, 1090 Elkton Road, Newark, Delaware 19711, USA (W.S.); LL Silver Consulting, 3403 Park Place, Springfield, New Jersey 07081, USA (L.L.S.).

*These authors contributed equally to the work.

‡Deceased.

Platensimycin shows potent, broad-spectrum Gram-positive activity *in vitro* (Table 1; methods described in ref. 14) and exhibits no cross-resistance to other key antibiotic-resistant bacteria including methicillin-resistant *S. aureus*, vancomycin-intermediate *S. aureus*, vancomycin-resistant enterococci, and linezolid-resistant and macrolide-resistant pathogens. Platensimycin showed antibacterial activity against efflux-negative *Escherichia coli* (*tolC*), but not against wild-type *E. coli*, indicating that efflux mechanisms, and not compound specificity, limit the effectiveness of platensimycin in *E. coli* and possibly other Gram-negative bacteria. Low mammalian cell toxicity (50% inhibitory concentration (IC_{50}) > 1,000 $\mu\text{g ml}^{-1}$ in a HeLa cytotoxicity assay) and lack of antifungal activity (more than 64 $\mu\text{g ml}^{-1}$) further suggests that platensimycin acts selectively. Platensimycin exhibited minimum inhibitory concentration (MIC) values of 0.5 and 1 $\mu\text{g ml}^{-1}$ against *S. aureus* and *S. pneumoniae*, respectively.

The efficacy of platensimycin *in vivo* was evaluated in a mouse model of a disseminated *S. aureus* infection. In this experiment, mice were cannulated in the jugular vein and drug was delivered with a continuous-delivery pump. The compound showed excellent efficacy with a 10^4 – 10^5 fold decrease (4–5 log reduction) of *S. aureus* infection over a 24-h treatment period (Fig. 1b) at 100 and 150 $\mu\text{g h}^{-1}$ with no indication of toxicity. Monitoring of plasma concentration during treatment provided a dose-dependent exposure level of about 0.17 $\mu\text{g ml}^{-1}$ (compound administered at 25 $\mu\text{g h}^{-1}$), 0.32 $\mu\text{g ml}^{-1}$ (50 $\mu\text{g h}^{-1}$) and about 0.6 $\mu\text{g ml}^{-1}$ (100 $\mu\text{g h}^{-1}$). The efficacious dose of 100 $\mu\text{g h}^{-1}$ produced a plasma exposure level of 0.6 $\mu\text{g ml}^{-1}$, which is lower than the MIC *in vitro* (2 $\mu\text{g ml}^{-1}$) determined in the presence of serum (Table 1, MIC tested *in vitro* with serum to mimic the animal test conditions), indicating a possible minimal serum effect *in vivo*. No toxicity was observed in similar experiments in which mice received 300 or 250 $\mu\text{g h}^{-1}$ for 24 h, and in a longer study in which 50 $\mu\text{g h}^{-1}$ was administered for 9 days followed by 100 $\mu\text{g h}^{-1}$ for 8 days.

In whole-cell labelling experiments (Fig. 1c), platensimycin showed selective inhibition of lipid biosynthesis in *S. aureus* with an IC_{50} of 0.1 $\mu\text{g ml}^{-1}$. Platensimycin did not inhibit DNA, RNA, protein or cell wall biosynthesis at concentrations up to 500 $\mu\text{g ml}^{-1}$, thereby confirming its selectivity. Similar results were obtained with *S. pneumoniae*, in which the IC_{50} for inhibition of cellular lipid biosynthesis was 0.8 $\mu\text{g ml}^{-1}$. The correspondence between MICs and IC_{50} values for the inhibition of cellular lipid biosynthesis (Table 1) strongly indicates that the antibiotic kills bacteria entirely through the inhibition of fatty-acid biosynthesis.

Table 1 | Microbiological profiles and toxicity of platensimycin and linezolid

Organism and genotype	Platensimycin	Linezolid
Antibacterial activity (MIC, $\mu\text{g ml}^{-1}$)*		
<i>S. aureus</i> (MSSA)	0.5	4
<i>S. aureus</i> + serum	2	4
<i>S. aureus</i> (MRSA)	0.5	2
<i>S. aureus</i> (MRSA, macrolide ^R)	0.5	2
<i>S. aureus</i> (MRSA, linezolid ^R)	1	32
<i>S. aureus</i> (VISA, vancomycin ^I)	0.5	2
<i>Enterococcus faecalis</i> (macrolide ^R)	1	1
<i>Enterococcus faecium</i> (VRE)	0.1	2
<i>S. pneumoniae</i> †	1	1
<i>E. coli</i> (wild-type)	>64	>64
<i>E. coli</i> (<i>tolC</i>)	16	32
Toxicity ($\mu\text{g ml}^{-1}$)		
HeLa MTT (IC_{50})	>1,000	>100
<i>Candida albicans</i> (MIC)	>64	>64

* A concentration of 1 $\mu\text{g ml}^{-1}$ equals 2.27 μM for platensimycin and 2.96 μM for linezolid.

† Cells were inoculated at 10^5 colony-forming units followed by incubation overnight at 37 °C with a serial dilution of compounds in Todd-Hewitt broth.

Linezolid is a synthetically derived agent that has been in clinical use since 2000. MRSA, methicillin-resistant *S. aureus*; MSSA, methicillin-susceptible *S. aureus*; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide; VISA, vancomycin-intermediate *S. aureus*; VRE, vancomycin-resistant *Enterococcus*.

The target selectivity of platensimycin was first determined with *S. aureus* strains either expressing antisense FabF RNA (AS-*fabF*) or overexpressing FabF protein. Antisense expression decreases expression of the targeted protein, making the cells more sensitive to exogenous inhibitors of the targeted protein. Hence, antisense plates show larger inhibition zones than wild-type control plates¹⁵ (Supplementary Fig. S1a), whereas overexpression strains show the opposite effect (Supplementary Fig. S1b). In comparison with cerulenin, platensimycin showed a 200-fold greater potency and similar FabF target selectivity in both assays, indicating that FabF is the sole target of platensimycin for the inhibition of bacterial growth. This mechanism of action was further confirmed with a cell-free gel-elongation assay¹⁶ (Supplementary Fig. S2a). Potency analysis of platensimycin against FASII *S. aureus* enzymes in the gel-elongation assay provided an IC_{50} of 0.2 $\mu\text{g ml}^{-1}$ (Supplementary Fig. S2b). This value is consistent with the IC_{50} (0.1 $\mu\text{g ml}^{-1}$) obtained in the whole-cell labelling assay, indicating that platensimycin has good permeability through the cell membrane.

Single-enzyme catalytic assays showed that platensimycin is a potent inhibitor of FabF with an IC_{50} of 48 and 160 nM for *S. aureus* and *E. coli* enzymes, respectively (Supplementary Fig. S3a). In a similar catalytic assay, platensimycin showed weak inhibition of *S. aureus* FabH, with an IC_{50} of 67 μM (Supplementary Fig. S3b).

To demonstrate that platensimycin inhibits FabF by targeting the protein directly, we developed a direct binding assay employing a radioactive derivative of platensimycin (³H-dihydroplatensimycin; Supplementary Fig. S4a) and the purified, recombinant enzyme. Initial assays yielded binding affinities significantly less than those expected from the potency observed in other *in vitro* and *in vivo*

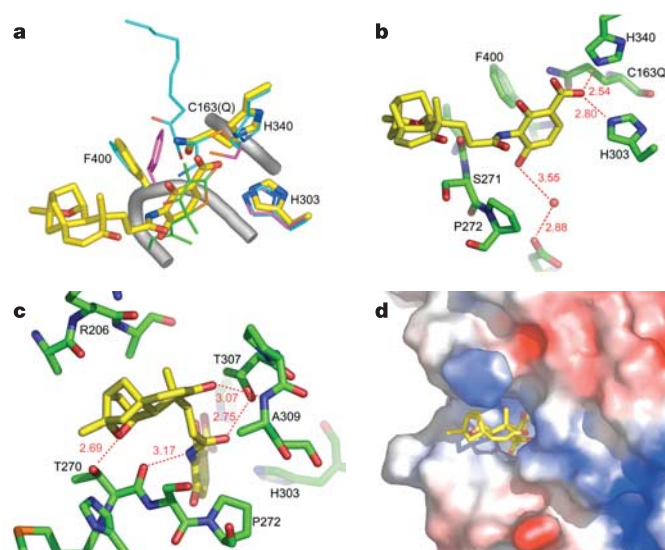


Figure 2 | Interactions of platensimycin with ecFabF(C163Q) and comparison with the apo structure. **a**, Superposition of platensimycin (yellow, thicker sticks) on ecFabF, with thiolaactomycin (green) and cerulenin (cyan) shown for reference. Side chains discussed in the text are labelled and coloured as described above. The side chains from apo ecFabF are coloured magenta. **b**, Interactions between the benzoic acid ring of platensimycin (yellow) and ecFabF(C163Q). Side chains from the protein discussed in the text are labelled and coloured green. **c**, Interactions of ecFabF with the amide linker and ketolide of platensimycin. The colour scheme is the same as in **b**. **d**, The solvent-accessible surface area of FabF, coloured according to electrostatic potential. Platensimycin is depicted as a stick figure and coloured yellow, and is shown to be partly exposed to solvent. Platensimycin buries 345 Å² of solvent-accessible surface area on ecFabF, as calculated with areaimol^{24,25}. Of that surface area, 122 Å² is a direct result of the ketolide portion of the molecule, highlighting its important contribution to platensimycin binding. Significant interatomic distances (in ångströms) are marked in **b** and **c** with red dashed lines and numbers.

assays. Changes in experimental conditions (for example pH or ionic strength) did not strongly increase inhibitor binding. Because the condensing enzymes catalyse a complex series of reactions using multiple substrates in a ping-pong mechanism¹⁷, we hypothesized that platensimycin might act by inhibiting a major species of the catalytic cycle, the covalent acyl-enzyme intermediate. We prepared a surrogate intermediate by adding lauroyl-CoA to mimic the natural acyl-ACP substrate, which results in the formation of a transiently stable covalent acyl-enzyme complex through the formation of a thioester with the cysteine (Cys 163) at the active site. Addition of this substrate increased the binding signal 19-fold relative to that obtained with the apoenzyme (Fig. 1d). This finding is consistent with the hypothesis that formation of the acyl-enzyme intermediate is essential for platensimycin binding. Titration studies of various inhibitors performed in the presence of lauroyl-CoA indicate a binding IC_{50} of 19 nM for platensimycin, 97 nM for dihydroplatensimycin and 1.1 μ M for thiolactomycin (Supplementary Fig. S5).

Attempts to obtain homogeneous preparations of acylated, wild-type FabF enzyme suitable for crystallographic studies were not successful, presumably as a result of the short half-life of the thioester intermediate. For this reason, various mutant proteins were generated in attempts to stabilize the acyl-enzyme intermediate. On the basis of previous literature reports, we mutated Cys 163 to serine in an effort to stabilize the acyl-enzyme intermediate¹⁸; however, the occupancies of the acyl chain and platensimycin in the crystals were too low to yield a high-quality structure. A similar mutation (K335A)¹⁹ also did not show any inhibitor binding because of unexpected, mutation-induced, conformational changes (Supplementary Information). Mutation of the active-site cysteine in an animal β -ketoacyl synthase domain to glutamine (C161Q) results in an enzyme with potent malonyl decarboxylase activities²⁰, possibly by sterically mimicking the formation of the acyl-enzyme intermediate. We generated the corresponding *E. coli* (ec)FabF(C163Q) mutant and found that this mutant enzyme shows a 50-fold increase in apparent binding of platensimycin compared with wild-type enzyme (Fig. 1d). The 2.6-fold increase over the signal observed with wild-type enzyme and lauroyl-CoA is probably due to the partial acylation of FabF with this non-natural substrate, whereas the C163Q mutation presumably pre-positions nearly all components of the FabF catalytic machinery in the acyl-enzyme conformation, resulting in more robust binding. This interpretation is supported by structural studies of the unliganded ecFabF(C163Q) protein (Supplementary Information), showing that this mutation results in a protein with all the characteristic features of the acyl-enzyme structure. In particular, the mutation induces a change in the side chain of Phe 400, converting it from a 'closed' conformation—one that blocks platensimycin binding in the wild-type apoenzyme—to an 'open' conformation in the mutant (see below; Fig. 2a, and Supplementary Fig. S6).

The 2.6-Å structure of ecFabF(C163Q) in complex with platensimycin shows that the antibiotic binds in the malonyl subsite of FabF (Fig. 2a), with its benzoic acid ring in roughly the same orientation as thiolactomycin²¹, where it is positioned to interact with important residues of the catalytic machinery¹⁹ (Fig. 2b). The carboxylate of the benzoic acid ring engages the active-site histidine residues (His 303 and His 340) that are part of the conserved His-His-Cys catalytic triad (Fig. 2b). The distances between the carboxylate and His 303 (2.5 Å) and His 340 (2.8 Å), together with previous reports that these histidines are involved in the decarboxylation reaction¹⁹ and may be protonated to stabilize negative charges during the catalytic cycle²², are all indicative of a strong interaction with a probable ionic component. The location and proximity of the carboxylate in platensimycin to the catalytic histidines further suggests that this interaction mimics the interaction between the natural substrate, malonate, and the enzyme.

The benzoic acid ring stacks edge-to-face against the open conformation of Phe 400 (for discussions see Supplementary Information),

the gatekeeper residue that separates the malonyl-binding and acyl-binding subsites. The hydroxyl group *para* to the acid makes weak water-mediated hydrogen bonds to protein residues in the extreme reaches of the malonyl subsite (Fig. 2b). The observation of the energetically favourable edge-to-face interaction²³ between the benzoic acid ring and Phe 400 reveals why platensimycin binds selectively to the acyl-enzyme intermediate: In the apoenzyme the closed conformation of Phe 400 would clash sterically with platensimycin's benzoic acid ring and could not make as favourable a stacking interaction with the inhibitor (Fig. 2a, b).

The amide-linked, pentacyclic, ketolide portion of platensimycin is positioned in the mouth of the active site and is partly exposed to solvent (Fig. 2c, d). The amide linker between the benzoic acid ring and the ketolide makes two critical hydrogen bonds to the protein. The carbonyl oxygen bonds to the side chain of Thr 307, and the amide nitrogen interacts with the backbone carbonyl of Thr 270. The ketolide itself makes numerous van der Waals interactions with the protein and exploits two direct hydrogen bonds that contribute to binding specificity. The enone carbonyl oxygen makes a hydrogen bond to the backbone amide of Ala 309, and the ether oxygen makes a hydrogen bond to the side-chain hydroxyl of Thr 270. The structural results described here are fully consistent with a model for FabF inhibition by platensimycin based on competition with malonyl-ACP for the malonyl subsite of the enzyme. The nature of the interactions between platensimycin and FabF indicates that resistance arising through the mutation of target protein is unlikely (Supplementary Information).

The novel chemical architecture and unique mode of action and the encouraging results *in vivo*, coupled with the biochemical and structural studies, emphasize that platensimycin provides a great opportunity for the development of critically needed antibiotics.

METHODS

Screening for natural product inhibitors. Fermentation broth derived from *Streptomyces platensis* grown in a liquid medium was extracted with an equal volume of acetone and filtered. Acetone was removed by blowing with nitrogen and the aqueous portion was tested on to FabF (target-based) two-plate assay¹⁵ followed by confirmation in a FASII assay¹⁶.

Isolation of platensimycin. The active acetone extract was charged on a reverse-phase Amberchrome resin and eluted with a linear gradient of aqueous methanol. Each fraction was tested in the FabF two-plate assay, and fractions that showed activity (larger zone on the antisense plate than on the control plate) were pooled and rechromatographed on a Sephadex LH 20 column in methanol. The resulting fractions with FabF activity were purified by reverse-phase C_{18} high-performance liquid chromatography with a linear gradient of aqueous acetonitrile (containing trifluoroacetic acid), yielding platensimycin (2 mg, titre 3 mg l⁻¹) as an amorphous powder. The structure was elucidated by two-dimensional NMR analysis and mass spectra, and confirmed by single-crystal X-ray analysis.

***In vivo* study.** Mice were challenged intraperitoneally with *S. aureus*. Platensimycin was infused continuously intravenously at 100 μ l h⁻¹ for 24 h. After treatment, mice were killed and kidneys removed for the evaluation of bacterial infection (see Supplementary Information for further details).

Whole-cell labelling. The whole-cell labelling assay has been described previously¹⁶.

FabF direct binding assay. The binding of His-tagged recombinant FabF and [³H]dihydroplatensimycin was measured in a scintillation proximity assay format. Details are described in Supplementary Information.

Crystallographic study. X-ray data collection and determination of structures are detailed in Supplementary Information.

Received 10 March; accepted 6 April 2006.

1. Singh, S. B. & Barrett, J. F. Empirical antibacterial drug discovery—foundation in natural products. *Biochem. Pharmacol.* **71**, 1006–1015 (2006).
2. Butler, M. S. & Buss, A. D. Natural products—the future scaffolds for novel antibiotics? *Biochem. Pharmacol.* **71**, 919–929 (2006).
3. Campbell, J. W. & Cronan, J. E. Bacterial fatty acid biosynthesis: Targets for antibacterial drug discovery. *Annu. Rev. Microbiol.* **55**, 305–332 (2001).
4. Zhang, Y.-M., Marrakchi, H., White, S. W. & Rock, C. O. The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. *J. Lipid Res.* **44**, 1–10 (2003).

5. Heath, R. J. & Rock, C. O. Fatty acid biosynthesis as a target for novel antibacterials. *Curr. Opin. Investig. Drugs* **5**, 146–153 (2004).
6. Smith, S., Witkowski, A. & Joshi, A. K. Structural and functional organization of the animal fatty acid synthase. *Prog. Lipid Res.* **42**, 289–317 (2003).
7. White, S. W., Zheng, J., Zhang, Y. M. & Rock, C. O. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem.* **74**, 791–831 (2005).
8. Revill, W. P., Bibb, M. J., Scheu, A. K., Kieser, H. J. & Hopwood, D. A. Beta-ketoacyl acyl carrier protein synthase III (FabH) is essential for fatty acid biosynthesis in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **183**, 3526–3530 (2001).
9. Lai, C.-Y. & Cronan, J. E. β -Ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J. Biol. Chem.* **278**, 51494–51503 (2003).
10. Tsay, J., Rock, C. & Jackowski, S. Overproduction of beta-ketoacyl-acyl carrier protein synthase I imparts thiolactomycin resistance to *Escherichia coli* K-12. *J. Bacteriol.* **174**, 508–513 (1992).
11. Schujman, G. E., Choi, K. H., Altabe, S., Rock, C. O. & de Mendoza, D. Response of *Bacillus subtilis* to cerulenin and acquisition of resistance. *J. Bacteriol.* **183**, 3032–3040 (2001).
12. Matsumae, A., Nomura, S. & Hata, T. Studies on cerulenin. IV. Biological characteristics of cerulenin. *J. Antibiot. (Tokyo)* **17**, 1–7 (1964).
13. Noto, T., Miyakawa, S., Oishi, H., Endo, H. & Okazaki, H. Thiolactomycin, a new antibiotic. III. *In vitro* antibacterial activity. *J. Antibiot. (Tokyo)* **35**, 401–410 (1982).
14. Wang, J. *et al.* Discovery of a small molecule that inhibits cell division by blocking FtsZ, a novel therapeutic target of antibiotics. *J. Biol. Chem.* **278**, 44424–44428 (2003).
15. Young, K. *et al.* Discovery of FabH/FabF inhibitors from natural products. *Antimicrob. Agents Chemother.* **50**, 519–526 (2006).
16. Kodali, S. *et al.* Determination of selectivity and efficacy of fatty acid synthesis inhibitors. *J. Biol. Chem.* **280**, 1669–1677 (2005).
17. Heath, R. J. & Rock, C. O. The Claisen condensation in biology. *Nat. Prod. Rep.* **19**, 581–596 (2002).
18. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M. & Larsen, S. Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure (Camb)* **9**, 233–243 (2001).
19. McGuire, K. A., Siggaard-Andersen, M., Bangera, M. G., Olsen, J. G. & von Wettstein-Knowles, P. β -Ketoacyl-[acyl carrier protein] synthase I of *Escherichia coli*: aspects of the condensation mechanism revealed by analyses of mutations in the active site pocket. *Biochemistry* **40**, 9836–9845 (2001).
20. Witkowski, A., Joshi, A. K., Lindqvist, Y. & Smith, S. Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry* **38**, 11643–11650 (1999).
21. Price, A. C. *et al.* Inhibition of beta-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin. Structure and mechanism. *J. Biol. Chem.* **276**, 6551–6559 (2001).
22. Huang, W. *et al.* Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J.* **17**, 1183–1191 (1998).
23. Hunter, C. A., Singh, J. & Thornton, J. M. Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **218**, 837–846 (1991).
24. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
25. Collaborative Computational Project No. 4, The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Use of the IMCA-CAT beamline 17-ID (or 17-BM) at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with the Center for Advanced Radiation Sources at the University of Chicago. We thank the staff of IMCA-CAT for their able assistance in data collection. This paper is dedicated to the late John Barrett for his strong contributions to antibiotic research.

Author Contributions J.W. and S.B.S. led the discovery, microbiological, biochemical characterization and natural products chemistry. S.M.S. coordinated X-ray crystallographic analysis and direct binding assay development.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. Coordinates and structure factors for all structures presented in this paper have been deposited with the PDB under accession numbers 2GFV, 2GFW, 2GFX, and 2GFY. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W. (jun_wang2@merck.com), S.M.S. (stephen_soisson@merck.com) or S.B.S. (sheo_singh@merck.com).

LETTERS

A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth

Seyun Kim¹, Pauline Wong¹ & Pierre A. Coulombe^{1,2}

Cell growth, an increase in mass and size, is a highly regulated cellular event. The Akt/mTOR (mammalian target of rapamycin) signalling pathway has a central role in the control of protein synthesis and thus the growth of cells, tissues and organisms¹. A striking example of a physiological context requiring rapid cell growth is tissue repair in response to injury². Here we show that keratin 17, an intermediate filament protein rapidly induced in wounded stratified epithelia³, regulates cell growth through binding to the adaptor protein 14-3-3 σ . Mouse skin keratinocytes lacking keratin 17 (ref. 4) show depressed protein translation and are of smaller size, correlating with decreased Akt/mTOR signalling activity. Other signalling kinases have normal activity, pointing to the specificity of this defect. Two amino acid residues located in the amino-terminal head domain of keratin 17 are required for the serum-dependent relocalization of 14-3-3 σ from the nucleus to the cytoplasm, and for the concomitant stimulation of mTOR activity and cell growth. These findings reveal a new and unexpected role for the intermediate filament cytoskeleton in influencing cell growth and size by regulating protein synthesis.

Tissue injury triggers a homeostatic response in which cellular activation, growth and migration are coordinated to achieve repair, defined as the re-establishment of vital functions². The mechanisms responsible for integration of the various aspects of this complex response are unknown. By virtue of their abundance, properties and context-dependent regulation, cytoplasmic intermediate filaments are ideally suited to participate in this integration. These 10–12-nm wide, highly flexible, nonpolar cytoskeletal fibres are formed by a large group of diverse proteins that are differentially distributed and dynamically regulated throughout the body⁵. After injury to skin, wound-proximal epithelial cells increase in size, decrease in adhesiveness and polarize, concomitant with changes in protein regulation and gene expression^{2,6}. Among the keratin intermediate filament genes rapidly upregulated in wound-activated skin epithelial cells are *K6* (also known as *Krt6*), *K16* (also known as *Krt16*) and *K17* (also known as *Krt17*) (refs 2, 3). A similar phenomenon, involving distinct intermediate filament genes, occurs in other tissues⁷. Such wound-induced alterations in intermediate filament proteins probably alter cellular viscoelastic properties in a manner that optimizes tissue repair⁸. In light of the large number of signalling proteins already known to interact with intermediate filaments^{5,8}, these cytoskeletal fibres are well-placed to fulfill additional, non-mechanical functions during tissue repair and homeostasis.

K17-null (*K17*^{-/-}) mouse embryos⁴ show a delay in the closure of surface ectoderm wounds, but *K6a*^{-/-}; *K6b*^{-/-} double-knockout embryos do not⁹. To understand this phenotype, we analysed the surface ectoderm in embryonic day (E)11.5 mouse embryos. In wild-type embryos, wound-proximal epithelial cells upregulate *K17*⁹ and become significantly enlarged compared to intact tissue (Fig. 1a, b). This hypertrophic response is markedly blunted in *K17*^{-/-} embryos (Fig. 1a, b) but not in *K6a*^{-/-}; *K6b*^{-/-} embryos (data not shown).

Cellular hypertrophy, along with *K6*, *K16* and *K17* upregulation, also occurs in primary cultures of keratinocytes. Cultured *K17*^{-/-} keratinocytes are smaller than wild-type controls (Fig. 1c, d). Three-dimensional reconstruction of *K17*^{-/-} cells did not reveal changes in cell shape or thickness (data not shown), nor did the absence of *K17* significantly affect cell-matrix adhesion or cell cycling (Supplementary Fig. 1b, c). Thus, we suggest that *K17* may directly affect keratinocyte growth during wound repair and in primary culture.

De novo protein synthesis is essential for cell growth. In response to growth factors (Supplementary Fig. 1a), activation of phosphoinositide 3-OH kinase (PI(3)K) and Akt antagonizes the inhibitory

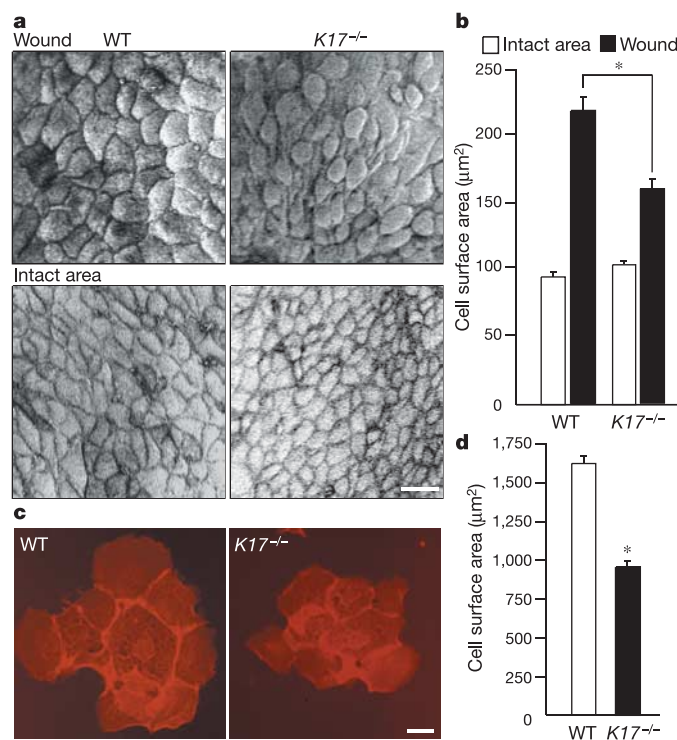


Figure 1 | *K17*^{-/-} keratinocytes are smaller than wild-type cells in the embryonic ectoderm and in primary cultures. **a**, The surface ectoderm of wounded E11.5 mouse embryos was analysed by scanning electron microscopy. Wounded area refers to tissue located within 100 μm of the wound edge, and intact area refers to the upper trunk of embryos. WT, wild type. **b**, Cell surface area measurements performed on preparations shown in **a**. **c**, p120-catenin staining was used to define the outer boundary of cultured keratinocytes³⁰. **d**, Cell surface area measurements of cells shown in **c**. For **b** and **d**, *n* > 100 cells for each genotype; data show mean ± s.e.m. Asterisk indicates *P* < 0.01 (Student's *t*-test). Scale bars in **a** and **c**, 20 μm.

¹Departments of Biological Chemistry and ²Dermatology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

activity of the tuberous sclerosis complex (TSC1/TSC2) on mTOR, a serine/threonine kinase with activity important to cell growth^{1,10}. TOR-mediated phosphorylation of p70 ribosomal S6 kinase (S6K1) and eukaryotic translation initiation factor 4E-binding protein 1 (4EBP1) stimulates protein synthesis^{10,11}. We therefore tested whether lack of K17 alters bulk protein translation and the activity of key regulators. Compared to wild-type cells, cultured *K17*^{-/-} keratinocytes have a 15% slower rate of amino acid incorporation into newly synthesized proteins (Fig. 2a), and a 73% slower rate of peptide chain elongation (Fig. 2b). mTOR and Akt activity are also significantly decreased in *K17*^{-/-} cells compared to wild-type cells (Fig. 2c and Supplementary Fig. 1d), whereas PI(3)K activity is unaffected (Fig. 2d), suggesting that K17 acts downstream of PI(3)K activation in this pathway (see below). Other kinases, including Erk1/2, p38 MAPK and JNK, have normal activity levels in *K17*^{-/-} cells (Fig. 2d), which do not show any signs of cellular stress¹² (Supplementary Fig. 1e). Autophagy is a lysosomal degradation process negatively regulated by mTOR¹³. *K17*^{-/-} keratinocytes show increase levels of the processed form of microtubule-associated protein-1 light chain-3 (LC3) (Fig. 2e), an autophagosome marker¹⁴, as well as the appearance of numerous autophagic vacuoles (Supplementary Fig. 2a). We conclude that the protein synthesis and growth defects in *K17*^{-/-} cells are specifically associated with down-regulation of the Akt/mTOR pathway.

We performed additional experiments to characterize *K17*^{-/-} keratinocytes. Treatment with rapamycin, a potent mTOR inhibitor¹⁰, reduces the rate of translation by 22% in wild-type cells and to a lesser extent (7%) in *K17*^{-/-} cells, as expected (Supplementary Fig. 1f). Upon serum removal, protein synthesis reaches the same low rate in *K17*^{-/-} and wild-type cells (Supplementary Fig. 2b). Serum re-addition causes dramatic stimulation of protein synthesis over 60 min in wild-type keratinocytes. In comparison, the stimulatory

effect of serum on protein synthesis was consistently ~15% lower in *K17*^{-/-} keratinocytes (Supplementary Fig. 2b). This depression in translation is also reflected in Akt/mTOR signalling. In wild-type cells, Akt and mTOR are robustly activated upon serum stimulation over at least 150 min. *K17*^{-/-} cells show marked activation of Akt 60 min after serum addition, but not at 150 min, and the activity of mTOR remains low in *K17*^{-/-} keratinocytes after serum addition (Supplementary Fig. 2b). Although an effect on upstream events cannot be ruled out, these findings imply that loss of K17 primarily affects one or more steps downstream of Akt activation. Given that mTOR can activate Akt¹⁵ (probably as part of a feedback regulatory mechanism), reduced mTOR activation could account for the decreased steady-state activity of Akt in *K17*^{-/-} cells (Supplementary Fig. 1a).

To identify K17-binding proteins that might provide mechanistic insight into this growth defect, we set up a biochemical screen using a reversible chemical crosslinker in cultured skin keratinocytes. We identified 14-3-3 σ (also known as stratifin), an epithelial-specific isoform of 14-3-3 (ref. 16), as a specific K17-binding protein in keratinocytes (Supplementary Fig. 3a). 14-3-3 proteins regulate many cellular processes by altering the conformation, activity or subcellular localization of target proteins, which they bind in a Ser/Thr-phosphorylation-dependent manner^{17,18}. Other intermediate filament proteins also bind 14-3-3 proteins^{19,20}; for example, keratin 18 binding influences 14-3-3 distribution and partially

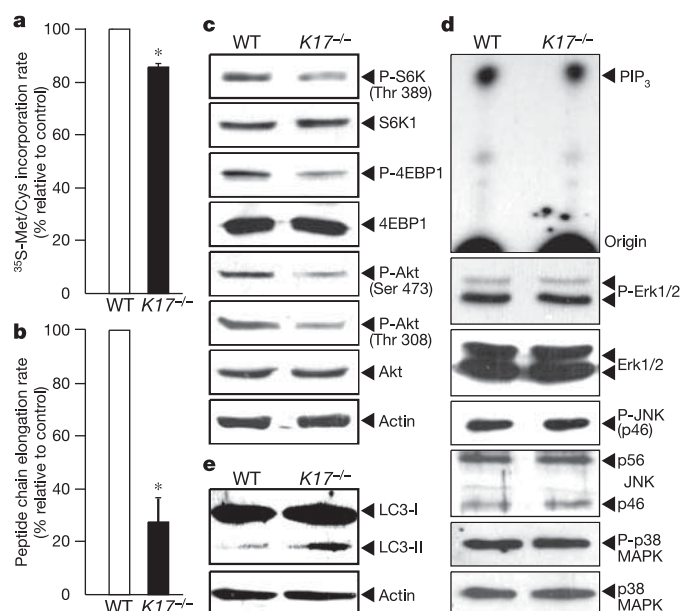


Figure 2 | Depressed protein translation and Akt/mTOR activity in *K17*^{-/-} keratinocytes. **a**, Total protein synthesis rate was measured by ³⁵S-methionine/cysteine incorporation into proteins. **b**, Peptide chain elongation rate was determined by measuring the ribosome-transit time²⁸. Data in **a**, **b** are mean $\pm$ s.d.; asterisk indicates $P < 0.01$ (Student's *t*-test). **c**, Lysates prepared from keratinocytes cultured in growth media were subjected to immunoblot analysis for phosphorylated (P) and non-phosphorylated proteins. **d**, PI(3)K and MAPK activity was assessed by kinase assay and immunoblot analysis, respectively. **e**, Immunoblot analysis for LC3-II, an autophagosome marker. In **c**–**e**, proteins being detected are identified to the right of each blot.

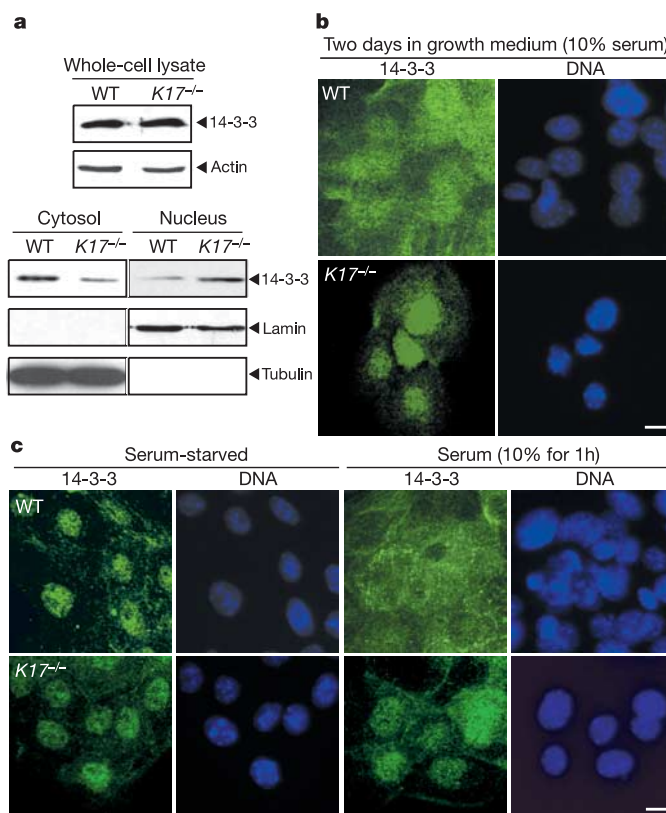


Figure 3 | Changes in subcellular localization of 14-3-3 proteins in *K17*^{-/-} keratinocytes. **a**, Whole-cell lysates, nuclear fractions and cytosolic fractions were prepared from keratinocytes cultured in growth media and subjected to immunoblot analysis using an anti-total-14-3-3 antibody. Actin is used as a loading control. Tubulin (cytosol) and lamin (nucleus) are used as controls for fractionation and loading. **b**, Immunostaining and confocal microscopy were used to assess the subcellular distribution of 14-3-3 proteins (green) in keratinocytes cultured in growth medium. The vital dye Hoechst was used to identify nuclei (blue). **c**, Cells were serum-starved overnight and then stimulated with 10% FBS for 1 h. 14-3-3 protein localization was determined as in **b**. Scale bars in **b**, **c**, 10 μ m.

affects mitotic progression in mouse liver hepatocytes²¹. We find that 14-3-3 σ colocalizes with K17 in the cytoplasm of skin keratinocytes, where their interaction is phosphorylation-dependent (Supplementary Fig. 3b–d). The distribution of K17 and 14-3-3 σ overlaps within hair follicles of intact skin tissue and, notably, in wounded epithelial tissues (Supplementary Fig. 3e). Given this coincidence, along with previous reports proposing 14-3-3 as a positive regulator of the mTOR pathway^{22–24} (Supplementary Fig. 1a), we investigated a possible role for 14-3-3 σ in the protein synthesis defects seen in *K17*^{-/-} cells.

Total 14-3-3 protein levels are similar in wild-type and *K17*^{-/-} keratinocytes (Fig. 3a). In wild-type keratinocytes grown in serum-containing growth medium, 14-3-3 σ is found in both the cytoplasm and nucleus, but it is found predominantly in the nucleus of *K17*^{-/-} cells (Fig. 3a, b). We reasoned that serum, or its physiological equivalent *in vivo*, would probably influence K17 phosphorylation and hence the partitioning of 14-3-3 σ in wild-type keratinocytes. Indeed, 14-3-3 proteins redistribute to the nucleus after serum withdrawal in wild-type cultures, and serum treatment causes

them to shift back to the cytoplasm (Fig. 3c). This correlates with increases in K17 phosphorylation and 14-3-3 binding (Supplementary Fig. 3c). In contrast, 14-3-3 remains preferentially localized to the nucleus in serum-stimulated *K17*^{-/-} keratinocytes (Fig. 3c), as reported for 14-3-3 mutants unable to bind their partners¹⁸. Such findings suggest a model (Supplementary Fig. 1a) in which K17 phosphorylation is required for the retention of 14-3-3 in the cytoplasm upon serum stimulation.

Identifying the amino acid residue(s) mediating 14-3-3 binding to K17 should help to define the role of this interaction with regards to protein synthesis. Two consensus 14-3-3-binding sites, threonine 9 (Thr9) and serine 44 (Ser44), are present in the non-helical head domain of K17 (Supplementary Fig. 4a). We mutated these sites to alanine residues and transfected the mutant DNA constructs into BHK fibroblasts. Relative to wild-type K17, T9A and S44A single mutants and the double mutant T9A/S44A show significantly reduced 14-3-3 binding (Supplementary Fig. 4b). The similar loss of binding observed for all mutants implies that Thr9 and Ser44 both contribute to binding to 14-3-3 (ref. 25).

To test for phenotypic rescue, we transiently transfected *K17*^{-/-} keratinocytes with wild-type K17, K17(T9A/S44A) or green fluorescent protein (GFP) as a control. Expression of wild-type K17 increases cytoplasmic 14-3-3 levels (Fig. 4a, b) and Akt/mTOR activity (Fig. 4c), increases the rate of protein synthesis by 30% (Fig. 4d) and results in an increased surface area of *K17*^{-/-} cells (Fig. 4e). In contrast, although the T9A/S44A mutant becomes readily incorporated into the keratin network, 14-3-3 remains in the nucleus of transfected cells (Fig. 4a, b), and there is no rescue effect on Akt/mTOR activity, translation or cell size (Fig. 4c–e). Moreover, transfection of 14-3-3 σ , either alone or together with the T9A/S44A mutant, causes a 20% increase in translation rate in *K17*^{-/-} cells, which corroborates the importance of cytosolic 14-3-3 protein levels (Supplementary Fig. 4c). Overexpression of wild-type K17 in wild-type cells has a similar stimulatory effect, but K17(T9A/S44A) overexpression in wild-type cells does not (Supplementary Fig. 4d). We thus suggest that Thr9 and/or Ser44 are required for the physiologically important role of K17 in stimulating Akt/mTOR signalling, protein synthesis and cell growth (Supplementary Fig. 1a).

Given that intermediate filaments are often modulated in wound-activated cells in several tissues⁷, the role we have identified for K17 probably also applies to other family members. In addition to tissue repair, K17 intermediate filaments could modulate keratinocyte growth during skin development and homeostasis as well as in diseases such as psoriasis and carcinoma²⁶. These findings add a new dimension to the biological significance of the profound cytoskeletal changes occurring in cells recruited to take part in tissue-remodelling events.

METHODS

Mice and cell cultures. *K17*^{-/-} mice were generated as previously described⁴. Skin keratinocytes from newborn mice were isolated and cultured as described²⁷. BHK21 fibroblasts were cultured in DMEM medium containing 10% fetal bovine serum (FBS).

Analysis of cell surface area in embryonic ectoderm and cultured keratinocytes. Using a Leo FEI scanning electron microscope (Zeiss), we analysed wound-proximal and intact ectoderm from E11.5 embryos that have previously been described⁹. Cell surface area was measured from photomicrographs of the surface ectoderm using MacBAS software. Two-day-old primary cultures of newborn mouse skin keratinocytes were immunostained for p120-catenin to reveal individual cell boundaries, and cell surface area was measured.

Analysis of cell cycling and plating efficiency. Relevant methods are described in the legend to Supplementary Fig. 1.

Rates of total protein synthesis and peptide chain elongation. Two-day-old primary cultures of newborn mouse skin keratinocytes were labelled with ³⁵S-methionine/cysteine (0.1 mCi ml⁻¹) in methionine- and cysteine-free DMEM for 30 min at 37°C. Proteins were precipitated using trichloroacetic acid (TCA) and the amount of incorporated radioactivity was measured by liquid scintillation. The rate of ³⁵S-Met/Cys incorporation per min per μ g of

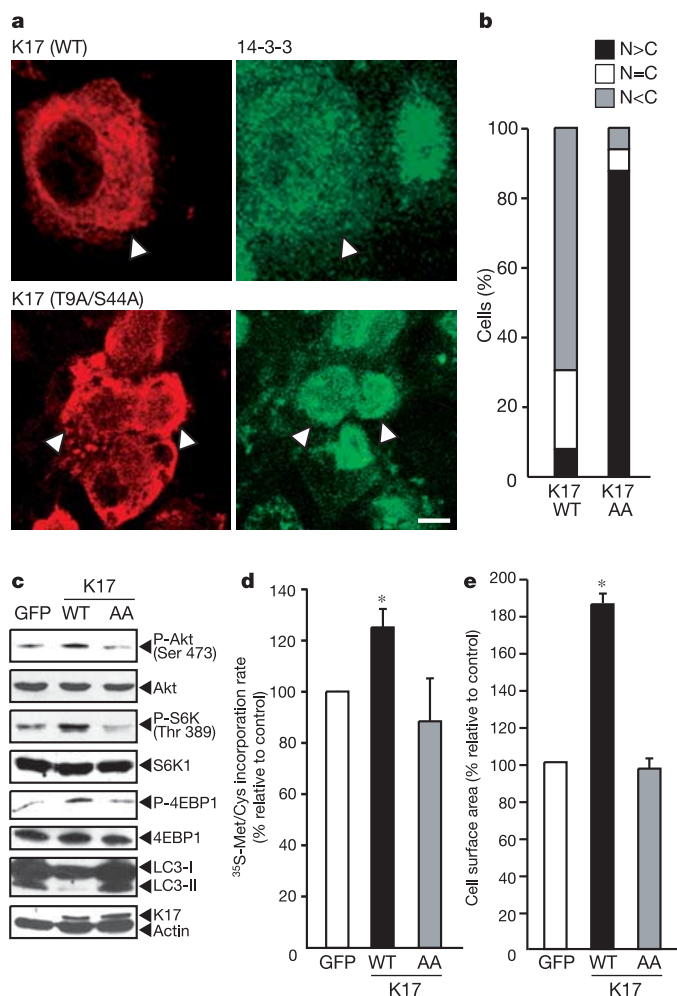


Figure 4 | Role of K17 in retaining 14-3-3 in the cytoplasm and promoting protein translation and cell growth. **a**, *K17*^{-/-} keratinocytes transfected with wild-type K17 or the K17(T9A/S44A) mutant (AA) were immunostained for K17 (red) and 14-3-3 (green), and analysed by confocal microscopy. Arrowheads point to transfected cells. Scale bar, 10 μ m. **b**, The 14-3-3 staining pattern was scored for localization to the nucleus (N) or cytoplasm (C) ($n > 50$ cells in each group). **c–e**, Akt/mTOR activity, translation rate and cell surface area of *K17*^{-/-} cells were determined in cells transfected as indicated. Data show mean $\pm$ s.d. (**d**), mean $\pm$ s.e.m. (**e**). Asterisk, $P < 0.01$ (Student's *t*-test). In **e**, $n > 30$ cells were analysed in each group.

protein was calculated. Translation elongation rate, defined as the inverse of the ribosome transit time, was measured from two-day-old primary cultures of newborn skin keratinocytes from wild-type and *K17^{-/-}* mice as previously described²⁸.

Antibodies. We used antibodies against 14-3-3 β (Santa Cruz Biotechnology), 14-3-3 σ (from B. Vogelstein), phospho-S473 and phospho-T308 Akt, Akt, phospho-T389 S6K1, S6K1, phospho-Erk, Erk, phospho-p38 MAPK, p38MAPK, phospho-JNK, JNK, phospho-T172 AMPK, AMPK (all from Cell Signalling), Hsp70, Hsp32 (StressGen) and LC3 (from T. Yoshimori).

³²P-Metabolic labelling, cell lysis, immunoblotting and immunoprecipitation. Metabolic labelling was done by incubating cells in phosphate-free DMEM containing 250 μ Ci ml⁻¹ ³²P-orthophosphate for 5 h. Lysates were prepared by incubating cells in 20 mM HEPES pH 7.5, 1% NP-40, 120 mM NaCl, 1 mM sodium vanadate, 2 mM NaF, 2.5 mM sodium pyrophosphate, 1 mM EDTA and protease inhibitors. Proteins in the lysates (8–10 μ g per lane) were resolved by SDS-PAGE and electroblotted. For western blot analysis, bound primary antibodies were detected by enhanced chemiluminescence (Pierce Biotechnology). Immunoprecipitation was also performed from these lysates using anti-K17 antibodies.

Nuclear and cytosolic fractionation. Cells were resuspended in hypotonic lysis buffer (20 mM Tris-HCl pH 7.5, 5 mM MgCl₂, 1 mM sodium vanadate, 10 mM NaF, 1.5 mM KCl, 0.1% NP-40 and protease inhibitors). Cells were incubated on ice for 10 min and homogenized using a 26-gauge syringe. Nuclei were pelleted by centrifugation at 800g for 10 min. The supernatant was centrifuged at 16,000g for 30 min to yield a crude cytoplasmic fraction.

PI(3)K activity assay. Cells were lysed in buffer containing 20 mM Tris-HCl pH 7.5, 137 mM NaCl, 1 mM MgCl₂, 1 mM CaCl₂, 10% glycerol, 1% NP-40, 150 μ M sodium vanadate and protease inhibitors. Immunoprecipitates were obtained using anti-phosphotyrosine or anti-PI(3)K p85 subunit antibodies. After washing three times with lysis buffer, twice with 0.5 M LiCl, 100 mM Tris-HCl pH 7.5, and twice with 10 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EDTA, beads were incubated in 50 μ l kinase buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 70 μ M phosphoinositide, 40 μ M ATP, 10 μ Ci ³²P-ATP) at 22 °C for 20 min. The reaction was stopped by adding 100 μ l MeOH:HCl (1:1 ratio), and the lipids were extracted with 200 μ l CHCl₃:MeOH (1:1 ratio). Phosphorylated lipids were spotted on a thin-layer chromatography (TLC) plate and separated with chloroform:MeOH:ammonium:water at a ratio of 60:47:2:11.3.

Transfection and cDNA constructs. For expression in mammalian cells, mouse *K17* cDNA was subcloned into pcDNA3.0 (Invitrogen) harbouring no tag sequence. Mutant *K17* coding sequences, in which alanine residues were introduced at Thr9 (T9A), Ser44 (S44A), and both Thr9 and Ser44, were generated by site-directed mutagenesis (Stratagene). For studies with BHK fibroblasts, *K17* cDNAs were co-transfected with an expression vector encoding mouse *K6a* (ref. 9). Plasmid pHR-14-3-3 σ , in which a cytomegalovirus promoter drives human 14-3-3 σ expression, was obtained from B. Vogelstein. Transient transfection was performed using Lipofectamine 2000 (Invitrogen).

Identification of 14-3-3 σ as a keratin-17-binding protein. We modified a protocol making use of a cell-permeable and reversible crosslinker, DSP (dithiobis[succinimidylpropionate])²⁹. To increase keratin solubilization, cell lysates were treated with 2% Empigen-BB for 30 min. A 30-kDa band present in *K17* immunoprecipitates was excised, digested with trypsin and analysed by matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Mascot (Matrix Science) was used to identify mouse protein sequences in the NCBI database.

Received 7 December 2005; accepted 17 February 2006.

- Kozma, S. C. & Thomas, G. Regulation of cell size in growth, development and human disease: PI3K, PKB and S6K. *Bioessays* **24**, 65–71 (2002).
- Martin, P. Wound healing—aiming for perfect skin regeneration. *Science* **276**, 75–81 (1997).
- Paladini, R. D., Takahashi, K., Bravo, N. S. & Coulombe, P. A. Onset of re-epithelialization after skin injury correlates with a reorganization of keratin filaments in wound edge keratinocytes: defining a potential role for keratin 16. *J. Cell Biol.* **132**, 381–397 (1996).
- McGowan, K. M. *et al.* Keratin 17 null mice exhibit age- and strain-dependent alopecia. *Genes Dev.* **16**, 1412–1422 (2002).
- Omary, M. B., Coulombe, P. A. & McLean, W. H. Intermediate filament proteins and their associated diseases. *N. Engl. J. Med.* **351**, 2087–2100 (2004).

- Bereiter-Hahn, J. in *Biology of the Integument-Vertebrates* (eds Bereiter-Hahn, J., Matoltsy, A. G. & Richards, K. S.) 443–463 (Springer-Verlag, New York, 1986).
- DePianto, D. & Coulombe, P. A. Intermediate filaments and tissue repair. *Exp. Cell Res.* **301**, 68–76 (2004).
- Coulombe, P. A. & Wong, P. Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds. *Nature Cell Biol.* **6**, 699–706 (2004).
- Mazzalupo, S., Wong, P., Martin, P. & Coulombe, P. A. Role for keratins 6 and 17 during wound closure in embryonic mouse skin. *Dev. Dyn.* **226**, 356–365 (2003).
- Hay, N. & Sonenberg, N. Upstream and downstream of mTOR. *Genes Dev.* **18**, 1926–1945 (2004).
- Fingar, D. C., Salama, S., Tsou, C., Harlow, E. & Blenis, J. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/elf4E. *Genes Dev.* **16**, 1472–1487 (2002).
- Hardie, D. G. The AMP-activated protein kinase pathway—new players upstream and downstream. *J. Cell Sci.* **117**, 5479–5487 (2004).
- Meijer, A. J. & Codogno, P. Regulation and role of autophagy in mammalian cells. *Int. J. Biochem. Cell Biol.* **36**, 2445–2462 (2004).
- Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membrane after processing. *EMBO J.* **19**, 5720–5728 (2000).
- Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).
- Prasad, G. L., Valverius, E. M., McDuffie, E. & Cooper, H. L. Complementary DNA cloning of a novel epithelial cell marker protein, HME1, that may be down-regulated in neoplastic mammary cells. *Cell Growth Differ.* **3**, 507–513 (1992).
- Fu, H., Subramanian, R. R. & Masters, S. C. 14-3-3 proteins: structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.* **40**, 617–647 (2000).
- Brunet, A. *et al.* 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. *J. Cell Biol.* **156**, 817–828 (2002).
- Liao, J. & Omary, M. B. 14-3-3 proteins associate with phosphorylated simple epithelial keratins during cell cycle progression and act as a solubility cofactor. *J. Cell Biol.* **133**, 345–357 (1996).
- Tzivion, G., Luo, Z. J. & Avruch, J. Calyculin A-induced vimentin phosphorylation sequesters 14-3-3 and displaces other 14-3-3 partners *in vivo*. *J. Biol. Chem.* **275**, 29772–29778 (2000).
- Ku, N. O., Michie, S., Resurreccion, E. Z., Broome, R. L. & Omary, M. B. Keratin binding to 14-3-3 proteins modulates keratin filaments and hepatocyte mitotic progression. *Proc. Natl Acad. Sci. USA* **99**, 4373–4378 (2002).
- Beck, T. & Hall, M. N. The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. *Nature* **402**, 689–692 (1999).
- Bertram, P. G., Zeng, C., Thorson, J., Shaw, A. S. & Zheng, X. F. The 14-3-3 proteins positively regulate rapamycin-sensitive signaling. *Curr. Biol.* **8**, 1259–1267 (1998).
- Li, Y., Inoki, K., Yeung, R. & Guan, K. L. Regulation of TSC2 by 14-3-3 binding. *J. Biol. Chem.* **277**, 44593–44596 (2002).
- Yaffe, M. B. *et al.* The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell* **91**, 961–971 (1997).
- Coulombe, P. A., Tong, X., Mazzalupo, S., Wang, Z. & Wong, P. Great promises yet to be fulfilled: defining keratin intermediate filament function *in vivo*. *Eur. J. Cell Biol.* **83**, 735–746 (2004).
- Bernot, K. M., Coulombe, P. A. & Wong, P. Skin: an ideal model system to study keratin genes and proteins. *Methods Cell Biol.* **78**, 453–487 (2004).
- Redpath, N. T., Foulstone, E. J. & Proud, C. G. Regulation of translation elongation factor-2 by insulin via a rapamycin-sensitive signalling pathway. *EMBO J.* **15**, 2291–2297 (1996).
- Kim, D. H. *et al.* mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell* **110**, 163–175 (2002).
- Staddon, J. M., Smales, C., Schulze, C., Esch, F. S. & Rubin, L. L. p120, a p120-related protein (p100), and the cadherin/catenin complex. *J. Cell Biol.* **130**, 369–381 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Coulombe laboratory for support, B. Vogelstein for providing 14-3-3 σ antibody and cDNA, T. Yoshimori for providing anti-LC3 antibody, and C. Parent, S. Craig, J. Lorsch, D. Ginty, S. H. Cha, S. Kim, C. S. Lee and C. Moon for advice. This work was supported by a grant from the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to P.A.C.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.A.C. (coulombe@jhmi.edu).

LETTERS

Critical role for the p110 α phosphoinositide-3-OH kinase in growth and metabolic regulation

Lazaros C. Foukas¹, Marc Claret², Wayne Pearce¹, Klaus Okkenhaug^{1†}, Stephen Meek³, Emma Peskett¹, Sara Sancho⁴, Andrew J. H. Smith³, Dominic J. Withers² & Bart Vanhaesebroeck^{1,5}

The eight catalytic subunits of the mammalian phosphoinositide-3-OH kinase (PI(3)K) family form the backbone of an evolutionarily conserved signalling pathway; however, the roles of most PI(3)K isoforms in organismal physiology and disease are unknown. To delineate the role of p110 α , a ubiquitously expressed PI(3)K involved in tyrosine kinase and Ras signalling, here we generated mice carrying a knockin mutation (D933A) that abrogates p110 α kinase activity. Homozygosity for this kinase-dead p110 α led to embryonic lethality. Mice heterozygous for this mutation were viable and fertile, but displayed severely blunted signalling via insulin-receptor substrate (IRS) proteins, key mediators of insulin, insulin-like growth factor-1 and leptin action. Defective responsiveness to these hormones led to reduced somatic growth, hyperinsulinaemia, glucose intolerance, hyperphagia and increased adiposity in mice heterozygous for the D933A mutation. This signalling function of p110 α derives from its highly selective recruitment and activation to IRS signalling complexes compared to p110 β , the other broadly expressed PI(3)K isoform, which did not contribute to IRS-associated PI(3)K activity. p110 α was the principal IRS-associated PI(3)K in cancer cell lines. These findings demonstrate a critical role for p110 α in growth factor and metabolic signalling and also suggest an explanation for selective mutation or overexpression of p110 α in a variety of cancers^{1,2}.

Class IA PI(3)Ks consist of a catalytic subunit (p110 α , p110 β or p110 δ) in complex with one of five distinct regulatory subunits, collectively referred to as 'p85s' (ref. 3). The p85 SH2 domains recruit the cytosolic PI(3)Ks via tyrosine-phosphorylated protein complexes to the PI(3)K lipid substrates in the plasma membrane. p110 δ is mainly found in leukocytes, and p110 δ -null mice have a range of immunological defects³. In contrast, mice homozygous for knockout alleles of the broadly expressed p110 α (encoded by *Pik3ca*) or p110 β (encoded by *Pik3cb*) die as embryos^{4,5}, whereas heterozygote p110 α and p110 β knockout mice have no metabolic or growth phenotypes⁶. Therefore, the relative contributions of these PI(3)K isoforms to normal mammalian physiology remain unknown.

Deletion of PI(3)K genes in mice often alters expression of non-targeted PI(3)K family members³, precluding accurate phenotypic analysis. For example, p85 is overexpressed in homozygous p110 α knockout embryos⁴, and expression of both p110 α and p110 β is reduced in p85 α knockout mice⁷. We therefore created knockin mice (carrying a germline mutation in the DFG motif of the p110 α ATP binding site) that express a kinase-dead p110 α ^{D933A} protein (Supplementary Fig. S1). In contrast to gene deletion, the knockin approach preserves signalling complex stoichiometry (illustrated in Fig. 1a). In heterozygous knockin mice, 50% of receptor–p110 α

complexes will be uncoupled from p110 α activity, regardless of the stoichiometry of receptor to p85–p110 α complexes. This contrasts with heterozygous knockout mice in which the capacity of any given receptor to signal via p110 α will be reduced only if p110 α expression is limiting relative to the receptor. Hence, heterozygous mice carrying the knockin mutation D933A (hereafter called p110 α ^{D933A/WT} mice) reveal the consequence of 50% loss of function of p110 α activity in all tissues where it is expressed.

p110 α ^{D933A/WT} mice were viable whereas homozygous p110 α ^{D933A/D933A} embryos were growth-retarded from embryonic day 9 (E9) and died around E10–11 (data not shown). p110 α , p110 β and p110 δ subunit expression in p110 α ^{D933A/WT} embryos and in liver, muscle and fat from p110 α ^{D933A/WT} mice was similar to that in wild-type embryos and tissues (Fig. 1b and data not shown). In contrast to homozygous p110 α knockout embryos⁴, p85 was not overexpressed in p110 α ^{D933A/D933A} embryos (Fig. 1b). Selective immunoprecipitation of the p110 α ^{D933A} protein via its carboxy-terminal Myc epitope tag (Supplementary Fig. S1) confirmed that this protein had lost catalytic activity (Fig. 1c). PI(3)K activity in p110 α immunoprecipitates was reduced by 50% and no compensatory alterations in the activities of p110 β and p110 δ were observed (Fig. 1c). Critically, p110 α ^{D933A/WT} mice displayed a 50% reduction in overall class IA PI(3)K activity in liver, muscle and fat, as assessed by either PDGF-receptor phosphopeptide pull down (Fig. 1d) or pan-p85 immunoprecipitations (data not shown). These data indicate that p110 α accounts for most of the class IA PI(3)K activity in these tissues. Therefore, the p110 α ^{D933A/WT} mouse permitted the interrogation of the physiological effects of reduced p110 α signalling at the organismal level.

p110 α ^{D933A/WT} mice showed reduced body weight (Fig. 1e) and length (Supplementary Fig. S2a). These parameters were also reduced in p110 α ^{D933A/WT} embryos just before birth (Fig. 1f), indicating that growth retardation starts during embryonic development and suggesting abnormalities in insulin-like growth factor-1 (IGF-1) receptor signalling⁸. Impaired IGF-1-stimulated proliferation of p110 α ^{D933A/WT} embryonic fibroblasts confirmed such a defect (Fig. 1g). The smaller size of adult p110 α ^{D933A/WT} mice was due to reduced lean mass (Supplementary Fig. S2b)—specifically, reduced skeletal muscle mass (Supplementary Fig. S2c). Major internal organ weight was not altered (Supplementary Fig. S3) and no gross pathological abnormalities were observed in tissues of 12-week-old p110 α ^{D933A/WT} mice (Supplementary Table 1). Although PI(3)K and downstream effectors such as PDK-1 and p70S6K have been implicated in growth regulation via alterations in cell size or number^{9,10}, the size of a range of cell types was unaffected in p110 α ^{D933A/WT} mice (data not shown).

¹Ludwig Institute for Cancer Research, 91 Riding House Street, London W1W 7BS, UK. ²Centre for Diabetes and Endocrinology, University College London, Rayne Institute, 5 University Street, London WC1E 6JJ, UK. ³Gene Targeting Laboratory, The Institute for Stem Cell Research, University of Edinburgh, West Mains Road, Edinburgh EH9 3JQ, UK. ⁴Department of Medicine, University of Fribourg, 14 Chemin du Musée, CH-1700, Switzerland. ⁵Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK. [†]Present address: Babraham Institute, Cambridge CB2 4AT, UK.

p110 $\alpha^{D933A/WT}$ mice displayed insulin resistance with increased insulin levels when fasted (Fig. 2a) or during a glucose tolerance test (Fig. 2b), and had an impaired hypoglycaemic response to insulin (Fig. 2c; Supplementary Fig. S4a). An increased pancreatic β -cell area also suggested β -cell compensation to insulin resistance (Supplementary Fig. S4b). Whereas fasting blood glucose levels were equivalent to those of wild-type mice (Fig. 2d), p110 $\alpha^{D933A/WT}$ mice displayed impaired glucose clearance (Fig. 2e; see also Supplementary Fig. S4c). These abnormalities were seen in both sexes and in pure C57BL/6 or mixed 129Sv-C57BL/6 genetic backgrounds (data not shown). These findings demonstrate that p110 α regulates insulin sensitivity and glucose metabolism *in vivo*.

Despite their smaller size, young and 1-yr-old p110 $\alpha^{D933A/WT}$ mice were hyperphagic (Fig. 2f and data not shown) and displayed increased adiposity (Fig. 2g) with enlarged white adipose cells (Supplementary Fig. S4d). p110 $\alpha^{D933A/WT}$ mice showed elevated levels of plasma leptin (Fig. 2h), suggesting either primary CNS leptin resistance or adiposity-induced leptin resistance caused by defective hypothalamic insulin signalling¹¹ or other potential impairments in PI(3)K-dependent events in the hypothalamic circuits, such as neuronal development and maturation¹².

The receptors for insulin, IGF-1 and leptin directly or indirectly recruit IRS adaptor molecules to transmit their signals, mainly via IRS-bound PI(3)K^{13–16}. There are four mammalian IRS proteins, of which IRS-1 and IRS-2 are widely expressed¹⁷. *Irs1*-null mice display growth retardation due to defective IGF-1 action, mild insulin resistance and glucose intolerance. In contrast, *Irs2* deletion results in diabetes due to insulin resistance and β -cell failure, and neuroendocrine dysfunction. p110 α inactivation did not affect the

expression levels of insulin receptor (IR) or IRS-1/2 under basal conditions (Fig. 3a). Insulin-stimulated tyrosine phosphorylation of IR and IRS and recruitment of p85 to IRS complexes were also equivalent in wild-type and p110 $\alpha^{D933A/WT}$ mice (Fig. 3a). In contrast, insulin-stimulated PI(3)K activity associated with IRS-1, IRS-2 or phosphotyrosine in skeletal muscle, liver and adipose tissue was reduced by approximately 50% in p110 $\alpha^{D933A/WT}$ mice (Fig. 3b and data not shown). Phosphorylation of Akt/protein kinase B (PKB), a key downstream target of PI(3)K implicated in insulin action and glucose homeostasis in mice *in vivo*¹⁸, was also reduced (Fig. 3c). These observations suggest that the defects in glucose homeostasis are due to reduced insulin-stimulated p110 α activity and diminished activation of effector pathways such as Akt/PKB in peripheral metabolic tissues.

Studies relying on the use of broad-spectrum pharmacological PI(3)K inhibition have suggested that food intake and adiposity are controlled by PI(3)K-dependent insulin and leptin signalling pathways in the CNS^{14–16}. IRS-1/2-associated PI(3)K activity in hypothalamic extracts of insulin- or leptin-treated p110 $\alpha^{D933A/WT}$ mice was severely reduced or absent (Fig. 3d), demonstrating that IRS-associated p110 α has either a direct role in insulin and leptin signalling in hypothalamic neurons or is required for the integrity of the hypothalamic circuits that respond to these hormones.

This predominant role of p110 α in insulin signalling *in vivo* contrasts with cell-based studies implicating p110 β in insulin-stimulated glucose uptake and cell proliferation^{19–22}. To examine how p110 α exerts its apparently selective role in IRS-mediated signalling, we performed quantitative immunoblot analysis of p110 α and p110 β expression, recruitment and activation in wild-type mice. p110 α

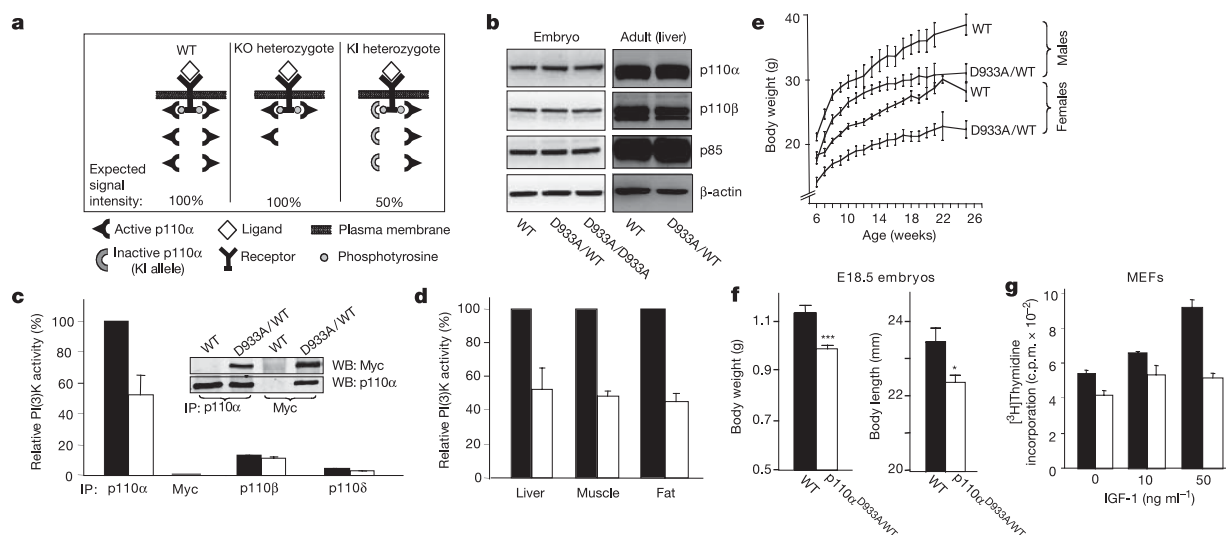


Figure 1 | Mechanism of action of p110 α D933A knockin strategy and effect on signalling and growth in p110 $\alpha^{D933A/D933A}$ embryos and p110 $\alpha^{D933A/WT}$ mice. **a**, Representation of the potential differential effect of p110 α knockout (KO, gene deletion) and knockin (KI, inactivation of gene product by point mutation) on the intensity of tyrosine kinase receptor-stimulated PI(3)K signalling. Binding of a growth factor to its cognate tyrosine kinase receptor results in tyrosine phosphorylation of the receptor and intracellular adaptor molecules (such as IRS proteins), leading to recruitment of class IA PI(3)Ks. If p110 α expression is not limiting, the remaining wild-type p110 α allele in cells heterozygous for the p110 α knockout allele will still give rise to sufficient p110 α protein to maintain a normal PI(3)K signalling output. In contrast, in mice heterozygous for the p110 α knockin allele, kinase-dead p110 α will become recruited to the receptor–adaptor complexes in the same way as wild-type p110 α (given that this occurs via p85) and reduce p110 α -dependent signalling by 50%, regardless of the stoichiometry of receptor to p85–p110 α . **b**, PI(3)K subunit expression. Homogenates of E9.5 embryos or adult liver were analysed by SDS–PAGE and immunoblotting using the indicated antibodies. **c**, PI(3)K

activity associated with distinct p110 isoforms. Homogenates of livers were immunoprecipitated using the indicated antibodies, followed by PI(3)K activity assay. The p110 α^{D933A} protein was immunoprecipitated via its C-terminal Myc epitope tag. Immunoprecipitation of p110 α^{D933A} was confirmed by immunoblotting (insert). **d**, PI(3)K activity associated with class IA PI(3)K regulatory subunits. Homogenates of the indicated tissues were absorbed onto Sepharose-immobilized phosphopeptide, which binds all class IA PI(3)K regulatory subunits, followed by *in vitro* lipid kinase assay. **e**, Reduced body weight of p110 $\alpha^{D933A/WT}$ mice. Body weight of a cohort of 11 male (wild type, $n = 6$; p110 $\alpha^{D933A/WT}$, $n = 5$) and 16 female ($n = 8$ per genotype) littermates was recorded over a period of 6 months. **f**, Weight and crown-to-rump length of E18.5 embryos isolated from pregnant females from ten litters (p110 $\alpha^{D933A/WT}$ ($n = 25$) and wild-type ($n = 15$) littermates). **g**, Impaired IGF-1-induced proliferation in MEFs from E13.5 p110 $\alpha^{D933A/WT}$ mice. For **c**, **d**, **f**, **g**, filled bars indicate wild-type mice and open bars indicate p110 $\alpha^{D933A/WT}$ mice. Error bars in all figures show standard error of the mean, except for Fig. 1e, which shows standard deviation. Asterisks indicate statistical significance (see Methods).

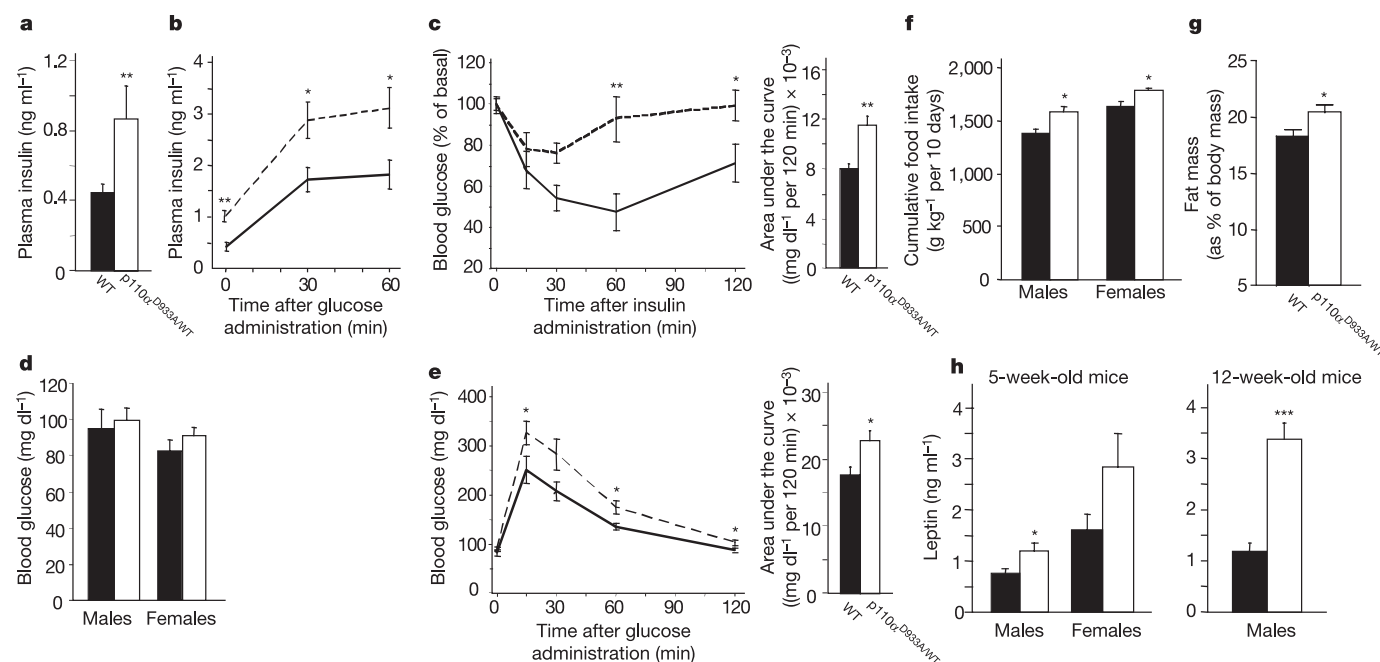


Figure 2 | Hyperinsulinaemia, impaired glucose and insulin tolerance, increased food intake, increased adiposity and hyperleptinaemia in p110α^{D933A/WT} mice. Solid lines/filled bars, wild-type mice; dashed lines/open bars, p110α^{D933A/WT} mice. **a**, Plasma insulin levels in fasted 8–9-week-old male mice ($n = 12$ per genotype). **b**, Insulin levels in the plasma of p110α^{D933A/WT} male mice and wild-type littermates ($n = 7$ per genotype) subjected to a glucose tolerance test. **c**, Insulin tolerance test in female p110α^{D933A/WT} ($n = 9$) and wild-type ($n = 9$) littermates. Calculated areas under the curves are shown. **d**, Blood glucose levels in fasted 8–9-week-old male and female mice ($n = 18$ per sex and genotype). **e**, Glucose tolerance

test in female p110α^{D933A/WT} ($n = 19$) mice and wild-type ($n = 17$) littermate mice. Calculated areas under the curves are shown. **f**, Food intake in a cohort of 12-week-old wild-type ($n = 6$ for males; $n = 7$ for females) and p110α^{D933A/WT} ($n = 5$ for males; $n = 6$ for females) littermates. **g**, Fat mass of p110α^{D933A/WT} mice ($n = 11$) and wild-type littermates ($n = 11$) was determined by DEXA analysis and expressed as per cent of total body mass. **h**, Plasma leptin levels in p110α^{D933A/WT} mice and wild-type littermates (5-week-old mice: males, $n = 16$ per genotype; females, $n = 9$ per genotype; 12-week-old mice: males, $n = 11$ per genotype). Asterisks indicate statistical significance (see Methods).

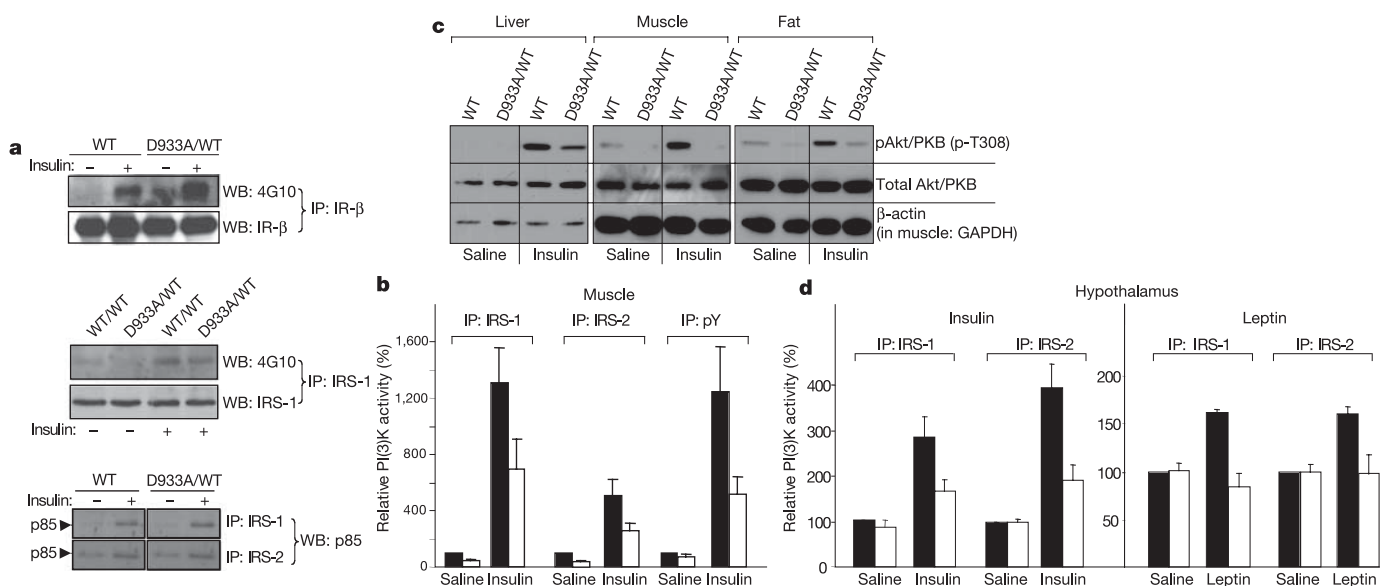


Figure 3 | Normal proximal insulin receptor signalling events but impaired PI(3)K pathway activation in p110α^{D933A/WT} mice. **a**, IR or IRS-1/2 were immunoprecipitated from livers of insulin-treated p110α^{D933A/WT} mice, immunoblotted with phosphotyrosine antibodies, followed by stripping and re-probing with antibodies to IR-β or IRS-1. For assessment of PI(3)K recruitment to IRS, IRS-1/2 immunoprecipitates were probed with antibodies to p85. **b**, **c**, p110α^{D933A/WT} mice were injected in the inferior

vena cava with saline or 0.1 mU g⁻¹ insulin. Liver, muscle (gastrocnemius) and perigenital fat pad homogenates were either immunoprecipitated using the indicated antibodies followed by PI(3)K lipid kinase assay (**b**) or analysed by SDS-PAGE and immunoblotting using the indicated antibodies (**c**). **d**, Defective hypothalamic IRS signalling in p110α^{D933A/WT} mice. IRS-1/2-associated PI(3)K activity in hypothalamic homogenates from mice injected with 5 U per mouse insulin or 1 μg g⁻¹ leptin.

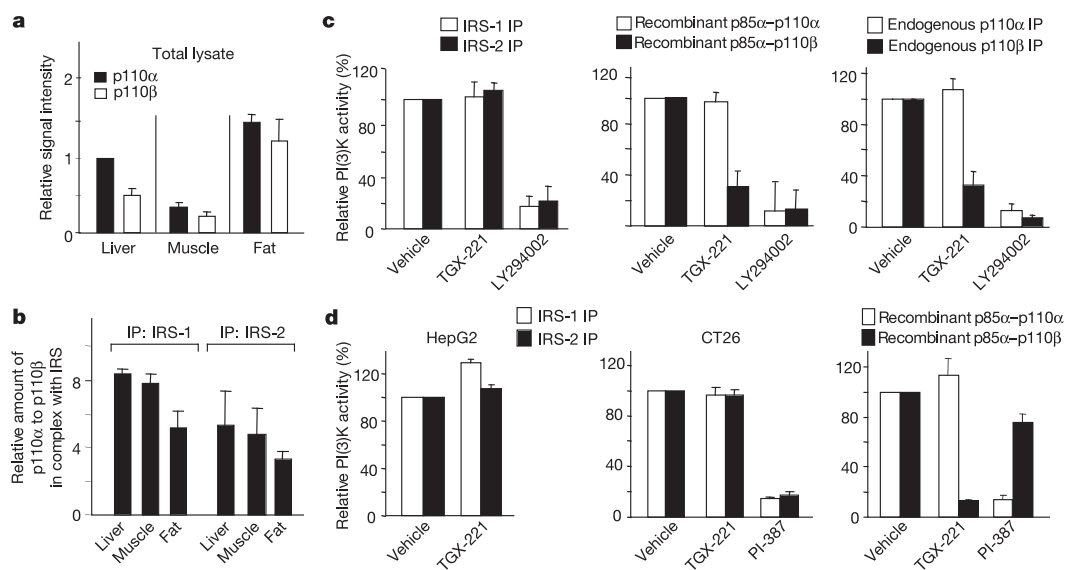


Figure 4 | Selective engagement of p110 α in IRS signalling. **a**, Relative expression of p110 α and p110 β in insulin-sensitive tissues of wild-type mice under basal conditions, as determined by two-colour immunoblot. The graph shows data acquired from six independent experiments. **b**, Association of p110 α and p110 β with IRS-1/2 in tissues of insulin-treated wild-type mice. IRS-1 or IRS-2 was immunoprecipitated from homogenates of organs isolated from wild-type mice injected with 5 U per mouse insulin, and associated p110 α and p110 β quantified as described in **a**. Data were pooled from two independent experiments. **c**, Effect of PI(3)K inhibitors on IRS-associated PI(3)K activity in liver. Left panel: sensitivity of IRS-1/2-associated PI(3)K activity to treatment with TGX-221 (p110 β -selective inhibitor) or LY294002 (pan-PI(3)K inhibitor). IRS-1/2 immunoprecipitates from liver homogenates of wild-type mice injected with

insulin were treated with vehicle or inhibitors (LY294002 at 5 μ M or TGX-221 at 100 nM), followed by an *in vitro* lipid kinase assay. Middle and right panels: effect of inhibitors on lipid kinase activity of recombinant p85 α -p110 α (middle panel) or p110 α and p110 β immunoprecipitates from liver homogenates (right panel). **d**, Sensitivity of IRS-1/2-associated PI(3)K activity to treatment with TGX-221 and PI-387 in cancer cell lines. IRS-1/2 immunoprecipitates from the indicated cell lines (HepG2 human liver hepatocarcinoma; CT26 murine colonic adenocarcinoma), which had been stimulated with 100 ng ml $^{-1}$ IGF-1 for 15 min, were treated with vehicle, TGX-221 (100 nM) or PI-387 (300 nM), followed by an *in vitro* PI(3)K lipid kinase assay. The effect of TGX-221 and PI-387 on recombinant p85 α -p110 α was tested in parallel (right panel).

levels were twofold higher than p110 β in liver and muscle, but similar in adipose tissue, which had the highest levels of p110 α and p110 β (Fig. 4a). Upon insulin stimulation, p110 α was significantly enriched compared with p110 β in IRS-1 and IRS-2 complexes in all three tissues (Fig. 4b). We next assessed the relative contribution of p110 α and p110 β to the insulin-stimulated PI(3)K lipid kinase activity associated with IRS proteins. TGX-221, a p110 β -selective ATP-competitive inhibitor²³, had no effect on IRS-associated lipid kinase activity at doses that inhibit recombinant or immunoprecipitated p110 β (Fig. 4c). In contrast, the pan-PI(3)K inhibitor LY294002 completely inhibited PI(3)K activity associated with IRS complexes (Fig. 4c). These data suggest a minimal contribution from p110 β towards IRS-associated lipid kinase activity. The intrinsically lower specific activity of p110 β compared to p110 α (ref. 24), or differences in the mechanism to fully activate p110 β (refs 25, 26), may underlie these findings. Taken together, our results demonstrate an unexpected specificity in IRS signalling, with selective recruitment and activation of p110 α to IRS complexes in metabolic tissues.

Selective functional association of p110 α with IRS complexes was also found in IGF-1-stimulated cancer cell lines. Indeed, IRS-associated PI(3)K activity was resistant to TGX-221 but sensitive to PI-387, an inhibitor with selectivity for p110 α over p110 β (Fig. 4d). IRS-mediated signalling is critical for IGF-1 receptor-driven cancer development and progression²⁷. The existence of a major signalling cassette composed of p110 α and IRS proteins in both growth factor and metabolic signalling may account for the observation that, of the eight distinct PI(3)K catalytic subunits in mammals, only p110 α is commonly mutated or overexpressed in cancer cells^{1,2}, possibly leading to selective activation of pathways involved in cell growth and metabolism.

Our knockin mouse gene targeting strategy has demonstrated that p110 α is a key intermediate in IGF-1, insulin and leptin signalling,

revealing a clearly defined role for p110 α at the organismal level. We have demonstrated specificity for p110 α in the relationship between PI(3)K catalytic subunits and IRS proteins. The observation that insulin resistance in p110 α ^{D933A/WT} mice does not progress to overt diabetes even at old age indicates a level of plasticity in metabolic control mechanisms after perturbation of the PI(3)K signalling axis. This suggests that small-molecule inhibitors of p110 α , currently under development as anticancer agents, may not lead to unmanageable metabolic disturbances^{28,29}.

METHODS

Mice. Mice were kept in individually ventilated cages and cared for according to UK Home Office regulations. Unless otherwise mentioned, the mice used were backcrossed to the C57BL/6 background for ten generations, and wild-type littermates were used as controls. Embryos were obtained from timed pregnant females. The day of presence of a copulation plug was considered day E0.5.

Reagents. A mouse monoclonal antibody specific for p110 α (clone U3A)³⁰ was used for two-colour immunoblotting. Anti-Akt/PKB (pThr308, pSer473 and total) and anti-Myc tag monoclonal antibody (clone 9B11) were from Cell Signalling Technologies. Antibodies to IRS-1, IRS-2 and phosphotyrosine (4G10) were from Upstate.

In vivo stimulation with insulin or leptin. Overnight-fasted mice (9–12 weeks old) were terminally anaesthetized by intraperitoneal injection of 150 mg kg $^{-1}$ sodium pentobarbital. Recombinant human insulin at doses varying from 0.01 mU g $^{-1}$ to 250 mU g $^{-1}$ or 5 U per mouse (Actrapid, Novo Nordisk) or recombinant mouse leptin (1 μ g g $^{-1}$; R&D Systems) were injected into the inferior vena cava, followed 5 min (for insulin) or 10 min (for leptin) later by harvesting and freezing of the tissues in liquid nitrogen.

Quantitative immunoblot analysis of p110 α and p110 β . Recombinant p110 α and p110 β in complex with p85 α were produced by baculovirus infection of Sf9 insect cells and purified by affinity chromatography using a Sepharose-immobilized phosphopeptide corresponding to Tyr 751 of PDGF receptor β . Purified PI(3)K proteins and a titration series of bovine serum albumin (BSA) were resolved by SDS-polyacrylamide gel electrophoresis (PAGE), followed by

Coomassie-blue staining and densitometric analysis of the gel using a Bio-Rad GS-800 densitometer, in order to determine the absolute amounts of p110 α and p110 β from the BSA protein standard curve. Equimolar amounts of p110 α and p110 β were then resolved by SDS-PAGE along with test samples (such as total protein from tissue homogenates), followed by simultaneous immunoblotting using antibodies specific for p110 α (mouse monoclonal; clone U3A) and p110 β (rabbit polyclonal; Santa-Cruz, S-19). These primary antibodies were detected using fluorescently labelled species-selective secondary antibodies (anti-mouse IRDye 800-conjugated (Rockland) and anti-rabbit Alexa-Fluor 680-conjugated (Molecular Probes)). Detection and quantification were performed using an Odyssey infrared scanner (LICOR) using the manufacturer's software. All signal intensities were normalized to those of the recombinant p110 α and p110 β standards in order to account for different avidity/affinity of the p110 α and p110 β antibodies.

Cell size analysis. Determination of diameter or volume of various cell types derived from p110 α ^{D933A/WT} or wild-type mice was performed with a CASY Model TT cell counter and analyser (Schärfe System).

MEF proliferation assay. E13.5 embryos were minced, dissociated with trypsin and cells allowed to adhere on gelatin-coated tissue culture dishes. For proliferation assays, early passage (P2–P4) mouse embryonic fibroblasts (MEFs) were incubated at 10⁵ cells per well in 96-well plates, deprived of fetal calf serum for 16 h followed by stimulation with or without IGF-1 in 200 μ l medium (DMEM, 10% FCS, penicillin/streptomycin). [³H]Thymidine was added for the last 4 h of a 24-h culture, followed by harvesting of the cells and quantification of incorporated [³H]thymidine by scintillation counting.

Metabolic studies and morphometric analysis. Measurements of body weight and length, glucose tolerance tests, plasma insulin and leptin level determination, and pancreatic morphometric analysis were performed as described elsewhere¹¹. Body composition was analysed by dual energy X-ray absorptiometry (DEXA) using a PIXImus 2 densitometer (Lunar). Insulin tolerance tests were performed on randomly fed mice by intraperitoneal injection of 0.75 U kg⁻¹ of recombinant human insulin, followed by measurement of blood glucose at various time points. For adipocyte morphometric analysis, epididymal fat pads were fixed in 10% neutral-buffered formalin, embedded in paraffin and cut into 8- μ m-thick sections. Adipose tissue sections were stained with haematoxylin and eosin, following standard protocols. In each mouse, adipocyte size was analysed in four sections (150 μ m apart) using Simple PCI software. At least 300 cells from each mouse were measured.

Food intake analysis. Mice were singly housed and allowed to acclimatize for a week before study. Food intake and body weight were measured for ten consecutive days. Results were expressed as cumulative food intake (grams of chow) per kg of body weight per 10 days. Experiments were performed on 12-week-old and 1-yr-old mice.

Statistical analysis. Values are presented as mean $\pm$ standard error of the mean. *P*-values were calculated using the non-parametric two-tailed Mann–Whitney *U*-test (comparisons with wild-type mice). *P*-values ≤ 0.05 were considered to be statistically significant (designated by a single asterisk; double asterisk, *P* ≤ 0.01 ; triple asterisk, *P* ≤ 0.001). The number of animals in each group is indicated by *n*.

Received 18 November 2005; accepted 2 March 2006.

Published online 12 April 2006.

- Shayesteh, L. *et al.* PIK3CA is implicated as an oncogene in ovarian cancer. *Nature Genet.* **21**, 99–102 (1999).
- Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
- Vanhaesebroeck, B., Ali, K., Bilancio, A., Geering, B. & Foulkes, L. C. Signalling by PI3K isoforms: Insights from gene-targeted mice. *Trends Biochem. Sci.* **30**, 194–204 (2005).
- Bi, L., Okabe, I., Bernard, D. J., Wynshaw-Boris, A. & Nussbaum, R. L. Proliferative defect and embryonic lethality in mice homozygous for a deletion in the p110 α subunit of phosphoinositide 3-kinase. *J. Biol. Chem.* **274**, 10963–10968 (1999).
- Bi, L., Okabe, I., Bernard, D. J. & Nussbaum, R. L. Early embryonic lethality in mice deficient in the p110 β catalytic subunit of PI 3-kinase. *Mamm. Genome* **13**, 169–172 (2002).
- Brachmann, S. M., Ueki, K., Engelman, J. A., Kahn, R. C. & Cantley, L. C. Phosphoinositide 3-kinase catalytic subunit deletion and regulatory subunit deletion have opposite effects on insulin sensitivity in mice. *Mol. Cell. Biol.* **25**, 1596–1607 (2005).
- Fruman, D. A. *et al.* Hypoglycemia, liver necrosis and perinatal death in mice lacking all isoforms of phosphoinositide 3-kinase p85 α . *Nature Genet.* **26**, 379–382 (2000).

- Dupont, J. & Holzenberger, M. Biology of insulin-like growth factors in development. *Birth Defects Res. C Embryo Today* **69**, 257–271 (2003).
- Lawlor, M. A. *et al.* Essential role of PDK1 in regulating cell size and development in mice. *EMBO J.* **21**, 3728–3738 (2002).
- Kozma, S. C. & Thomas, G. Regulation of cell size in growth development and human disease: PI3K, PKB and S6K. *Bioessays* **24**, 65–71 (2002).
- Choudhury, A. I. *et al.* The role of insulin receptor substrate 2 in hypothalamic and β cell function. *J. Clin. Invest.* **115**, 940–950 (2005).
- Bouret, S. G., Draper, S. J. & Simerly, R. B. Trophic action of leptin on hypothalamic neurons that regulate feeding. *Science* **304**, 108–110 (2004).
- White, M. F. Insulin signaling in health and disease. *Science* **302**, 1710–1711 (2003).
- Niswender, K. D. *et al.* Intracellular signalling: Key enzyme in leptin-induced anorexia. *Nature* **413**, 794–795 (2001).
- Niswender, K. D. *et al.* Insulin activation of phosphatidylinositol 3-kinase in the hypothalamic arcuate nucleus: A key mediator of insulin-induced anorexia. *Diabetes* **52**, 227–231 (2003).
- Xu, A. W. *et al.* PI3K integrates the action of insulin and leptin on hypothalamic neurons. *J. Clin. Invest.* **115**, 951–958 (2005).
- Withers, D. J. Insulin receptor substrate proteins and neuroendocrine function. *Biochem. Soc. Trans.* **29**, 525–529 (2001).
- Cho, H. *et al.* Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB β). *Science* **292**, 1728–1731 (2001).
- Wang, Q., Bilan, P. J., Tsakiridis, T., Hinek, A. & Klip, A. Actin filaments participate in the relocalization of phosphatidylinositol 3-kinase to glucose transporter-containing compartments and in the stimulation of glucose uptake in 3T3–L1 adipocytes. *Biochem. J.* **331**, 917–928 (1998).
- Asano, T. *et al.* p110 β is up-regulated during differentiation of 3T3–L1 cells and contributes to the highly insulin-responsive glucose transport activity. *J. Biol. Chem.* **275**, 17671–17676 (2000).
- Roche, S., Downward, J., Raynal, P. & Courtneidge, S. A. A function for phosphatidylinositol 3-kinase β (p85 α -p110 β) in fibroblasts during mitogenesis: Requirement for insulin- and lysophosphatidic acid-mediated signal transduction. *Mol. Cell. Biol.* **18**, 7119–7129 (1998).
- Hooshmand-Rad, R. *et al.* The PI 3-kinase isoforms p110 α and p110 β have differential roles in PDGF- and insulin-mediated signaling. *J. Cell Sci.* **113**, 207–214 (2000).
- Jackson, S. P. *et al.* PI 3-kinase p110 β : A new target for antithrombotic therapy. *Nature Med.* **11**, 507–514 (2005).
- Beeton, C. A., Chance, E. M., Foulkes, L. C. & Shepherd, P. R. Comparison of the kinetic properties of the lipid- and protein-kinase activities of the p110 α and p110 β catalytic subunits of class-Ia phosphoinositide 3-kinases. *Biochem. J.* **350**, 353–359 (2000).
- Kurosu, H. *et al.* Heterodimeric phosphoinositide 3-kinase consisting of p85 and p110 β is synergistically activated by the $\beta\gamma$ subunits of G proteins and phosphotyrosyl peptide. *J. Biol. Chem.* **272**, 24252–24256 (1997).
- Kubo, H., Hazeki, K., Takasuga, S. & Hazeki, O. Specific role of p85/p110 β in GTP-binding protein-mediated activation of Akt. *Biochem. J.* **392**, 607–614 (2005).
- Pollak, M. N., Schernhammer, E. S. & Hankinson, S. E. Insulin-like growth factors and neoplasia. *Nature Rev. Cancer* **4**, 505–518 (2004).
- Drees, B. E., Mills, G. B., Rommel, C. & Prestwich, G. D. Therapeutic potential of phosphoinositide 3-kinase inhibitors. *Expert Opin. Ther. Patents* **14**, 703–732 (2004).
- Amaravadi, R. & Thompson, C. B. The survival kinases Akt and Pim as potential pharmacological targets. *J. Clin. Invest.* **115**, 2618–2624 (2005).
- Klippel, A., Escobedo, J. A., Hirano, M. & Williams, L. T. The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity. *Mol. Cell. Biol.* **14**, 2675–2685 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Camps, T. Ruckle, C. Rommel (Serono Pharmaceutical Research Institute) and Plarmed for pharmacological agents; W. Fantl for antibody reagents; and T. Arnett and I. Orriss for providing access and help with the DEXA scanner. Personal support for L.F., M.C. and K.O. was provided in part by a European Union FP5 Programme grant, and for W.P. by a European Union grant. The main grant support for this project was by Diabetes UK and the Ludwig Institute for Cancer Research (to B.V.) and by the Wellcome Trust and MRC (to D.J.W.), with additional support of the Biotechnology and Biological Science Research Council (to B.V.). The ISCR Gene Targeting Laboratory was supported by the Biotechnology and Biological Science Research Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.V. (bartvanh@ludwig.ucl.ac.uk).

The mechanism of cell differentiation in *Bacillus subtilis*

Dagmar Iber¹, Joanna Clarkson², Michael D. Yudkin² & Iain D. Campbell³

Sporulation in *Bacillus subtilis* serves as a model for the development of two different cell types from a single cell¹. Although much information has been accumulated about the mechanisms that initiate the developmental programmes², important questions remain that can be answered only by quantitative analysis. Here we develop, with the help of existing and new experimental results, a mathematical model that reproduces published *in vitro* experiments and explains how the activation of the key transcription factor is regulated. The model identifies the difference in volume between the two cell types as the primary trigger for determining cell fate. It shows that this effect depends on the allosteric behaviour of a key protein kinase and on a low rate of dephosphorylation by the corresponding phosphatase; both predicted effects are confirmed experimentally.

A key question in developmental biology is how two sister cells with identical genomes follow different patterns of gene expression. The sporulation of *Bacillus subtilis* is an exceptionally favourable system for studying this question. Early in sporulation the cell divides asymmetrically into two compartments, a smaller prespore and a larger mother cell (Fig. 1a). These remain attached but initiate different programmes of transcription, using compartment-specific transcription factors (σ factors)¹. The first of these, σ^F , is crucial, because its correct activation is essential for regulating the other sigma factors of sporulation. σ^F is inactive in the predivisional cell and the mother cell, but is activated in the prespore. Its regulation depends on three other proteins, SpoIIAB (AB), SpoIIAA (AA) and

SpoIIE (IIE). AB is bifunctional: it can form a complex with, and thus inactivate, σ^F , and it can act as a specific protein kinase that phosphorylates AA to AA-P (Supplementary Fig. S1). AA-P is hydrolysed to AA by IIE, a serine phosphatase. AA interacts with the σ^F -AB-ATP complex, releasing σ^F , which then binds to the core RNA polymerase to form the active σ^F holoenzyme². The key reactions are summarized in Fig. 1b.

As σ^F holoenzyme appears in the prespore and not in the mother cell, the attack of AA on the σ^F -AB-ATP complex must be compartment-specific. Although this specificity is crucial to the success of sporulation, its mechanism is not clear. One suggestion is that, because the volume of the mother cell is at least fourfold that of the prespore, IIE, which is displayed only on the asymmetric septum, will have higher phosphatase activity in the smaller compartment, leading to a greater concentration of AA in the prespore than in the mother cell³. However, without a mathematical analysis it is unclear whether such a small difference could account for compartment-specific σ^F activation, and several other mechanisms have been proposed¹. Another quantitative difficulty involves the binding of σ^F to the core RNA polymerase. In the predivisional cell nearly all the RNA polymerase is in the form of σ^A holoenzyme; the formation of σ^F holoenzyme therefore requires σ^F to compete with σ^A . But the affinity of core RNA polymerase for σ^A is 25-fold that for σ^F , and the maximum concentration of σ^F is only about double that of σ^A (ref. 4), so how can σ^F holoenzyme be formed?

We have now constructed a mathematical model that provides the answer to these two questions and many others. The model was first developed and refined by fitting to numerous *in vitro* experiments (Supplementary Information) before being applied to the situation expected in the living cell. A first success of this rigorous analysis was the insight that AB is an allosteric protein. Positive cooperativity in AB-AA binding, predicted by the model, could be shown experimentally by a Scatchard plot (Fig. 2), obtained by measuring the binding of a non-phosphorylatable derivative of AA (AA-S58A)⁵ to

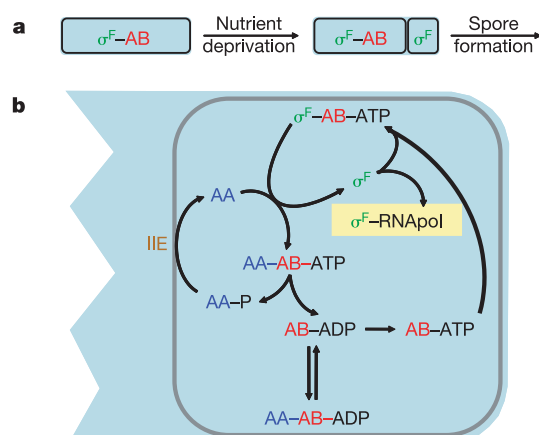


Figure 1 | Regulation of σ^F release during sporulation. **a**, Nutrient deprivation triggers asymmetric septation; the smaller, prespore compartment develops into a spore. **b**, Simplified regulatory network that controls σ^F activity in the prespore; for details see the text. RNAPol, RNA polymerase.

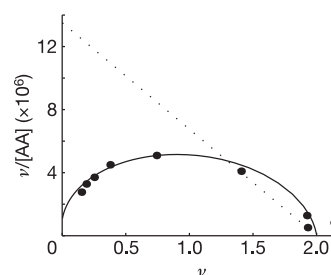


Figure 2 | SpoIIAB is an allosteric protein. Experimental (circles) and predicted Scatchard plot if AB is an allosteric (solid line) or non-allosteric (dotted line) protein; v is the fraction of AB that has bound AA.

¹Mathematical Institute, Centre for Mathematical Biology, University of Oxford, Oxford OX1 3LB, UK. ²Microbiology Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. ³Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.

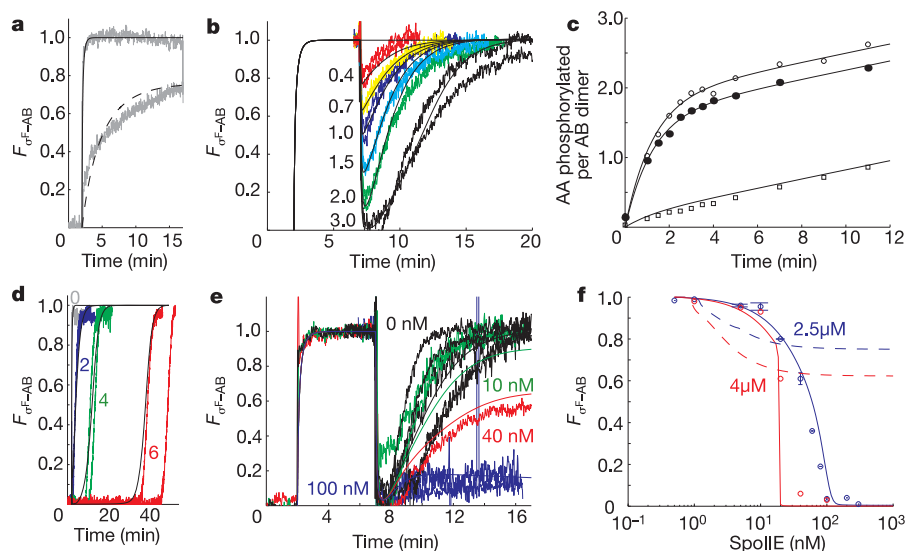


Figure 3 | Verification of the model with *in vitro* data. **a**, Measured (grey recorder output) and predicted (smooth lines) fraction F of σ^F in σ^F -AB complexes when 1.3 μM σ^F and 1 μM AB were mixed in the presence of 100 μM ATP (continuous line) or ADP (dashed line). **b**, Measured (recorder output) and predicted (smooth lines) σ^F release from AB-ATP on addition of AA. Black lines show the predictions for the indicated concentrations of AA. The measured values are for 0.4 μM (red), 0.7 μM (yellow), 1 μM (blue), 1.5 μM (cyan), 2 μM (green) and 3 μM (black), each carried out in duplicate. **c**, Measured (symbols) and predicted (lines) phosphorylation of AA by AB when AB is directly incubated with 40 μM AA and 100 μM ATP (open circles) or preincubated for 5 min with either 5 μM ADP (filled circles) or 5 μM ADP and 40 μM AA (open squares). The kinetics of phosphorylation of AA by AB was followed by measuring the uptake of radioactivity from [γ - ^{32}P]ATP into the protein over time, as described previously⁹. The experimental results are normalized by the factor 1.3 to take account of a

systematic error in the determination of protein concentrations (see Supplementary Information). **d**, Measured (recorder output) and predicted (black lines) σ^F -AB complex formation on mixing of 1 μM AB and 1.3 μM σ^F with no (grey), 2 μM (blue), 4 μM (green) or 6 μM (red) AA. Measurements were carried out in duplicate. **e**, Measured (recorder output) and predicted (smooth line) kinetics of IIE-dependent dissociation of σ^F -AB-ATP (1.3 μM σ^F , 1 μM AB dimer) by 2.5 μM AA, in the absence of IIE (five repeats as black lines) or with 10 nM (duplicates as green lines), 40 nM (red) or 100 nM (duplicates as blue lines) IIE. **f**, Measured (open circles) and predicted (lines) equilibrium fraction of σ^F complexed to AB as a function of different IIE concentrations, in the presence of 2.5 μM (blue) or 4 μM (red) AA, assuming that AB is allosteric (continuous line) or not allosteric with both AA and σ^F binding at high affinity (dashed). Error bars show s.d. from three experiments.

AB by surface plasmon resonance (SPR). We further find that the conformation of AB with high affinity for AA binds σ^F with low affinity and that the conformation with high affinity for σ^F binds AA with low affinity (Supplementary Information); this explains how AA binding to AB can result in the rapid dissociation of σ^F .

Because of AB allostery the different AB conformers and the interconversions between them have to be taken into account, and the simple description above grew into the more realistic regulatory scheme shown in Supplementary Fig. S2. Experimentally determined values were used for all the 25 rate constants (Supplementary Tables 1 and 2) that characterize the more than 150 reactions included in Supplementary Fig. S2.

Over the past few years numerous *in vitro* experiments have yielded much information about interactions between the components of the system². The use of fluorescent derivatives of AB and σ^F has been particularly fruitful in allowing the formation and dissociation of various complexes to be followed in real time⁶⁻⁹. Translation of the scheme of Supplementary Fig. S2 into a set of differential equations enabled us to simulate and predict the results of these experiments. As long as AB allostery was included, simulations and results agreed remarkably well for all experiments, for example those in which σ^F -AB complexes were made (Fig. 3a) and challenged (Fig. 3b) with different concentrations of AA. Simulations also gave excellent agreement with experiments in which AA was phosphorylated in the presence of AB, either with or without preincubation with ADP (Fig. 3c). Similarly, there was good agreement between the simulations and results of experiments in which AB and σ^F were mixed in the presence of different concentrations of AA (Fig. 3d). Here the model captured the increasing delay in the formation of the AB- σ^F complex with increasing concentrations of AA. This delay is due to the trapping of AA and AB-ADP in complexes,

which are important in the regulation of σ^F (refs 6, 10, 11). The model was highly successful in simulating the results of experiments in which σ^F -AB complexes were incubated with AA in the presence of different concentrations of IIE; the resemblance between experiment and simulation held for both the kinetics (Fig. 3e) and the equilibrium (Fig. 3f) of the process.

The success of the model in simulating *in vitro* experiments gave us confidence to apply it to the situation in the sporulating cell. Given that the model is relatively insensitive (robust) to small perturbations in most parameters (Supplementary Information, Supplementary Tables 1 and 2, and Supplementary Figs S2 and S12), parameter values derived from *in vitro* experiments could be used to model the physiological situation. The only exception was the rate of IIE phosphatase activity. The model predicted that this needed to be 75–150-fold smaller than measured *in vitro* to avoid septation-independent σ^F release (Fig. 4a). Earlier work had suggested that IIE activity might well be much lower in the cell than measured *in vitro*¹², at least partly because of the predominance of Mg^{2+} over Mn^{2+} (ref. 13). We tested this prediction of the model and found that for physiological salt concentrations (1 mM Mg^{2+} and 0.1 mM Mn^{2+} (ref. 14)) the phosphatase rate is indeed almost 150-fold lower than under *in vitro* conditions (2.5 mM Mn^{2+}) (Fig. 4b; Supplementary Information and Supplementary Figs S6–S8).

As a result of the accumulation of IIE on the asymmetrically positioned septum, the concentration of IIE increases in the smaller prespore³. To see whether such an increase is sufficient to account for the prespore-specific release of σ^F , we modelled the system with the known intracellular concentrations of the proteins (Supplementary Information and Supplementary Fig. S9) and then studied the effect of a fourfold increase in the IIE concentration (Supplementary Information). The simulation shows that little σ^F is released (Fig. 4c, dashed

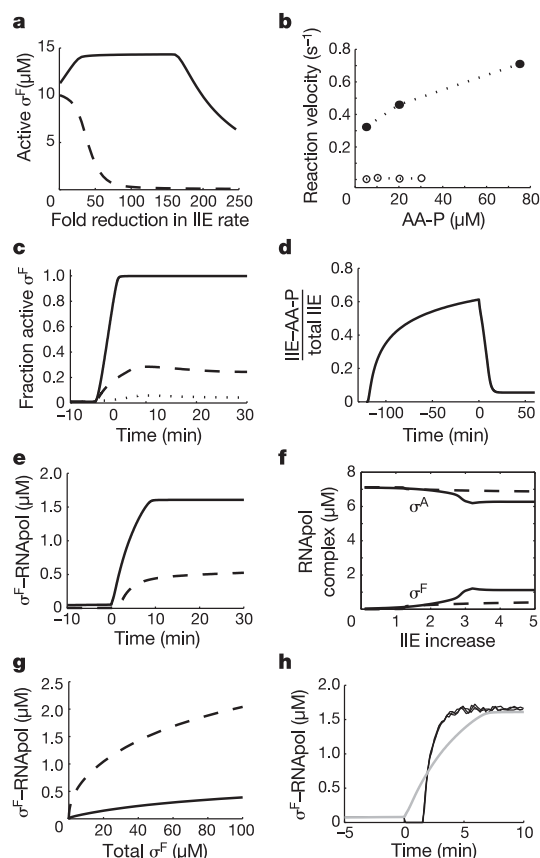


Figure 4 | Mechanism of prespore-specific σ^F -RNA-polymerase holoenzyme formation. **a**, Predicted concentration of AB-free (active) σ^F before (dashed line) and after (continuous line) septation, dependent on how much lower the IIE phosphatase rate is *in vivo* than under *in vitro* conditions. **b**, Velocity of the IIE reaction as a function of the concentration of its substrate AA-P in 2.5 mM Mn^{2+} (filled circles) or 1 mM Mg^{2+} and 0.1 mM Mn^{2+} (open circles). **c**, Predicted σ^F release in response to a fourfold increase in IIE alone (dashed line), AA-P alone (dotted line) or IIE and AA-P (solid line) imposed at $t = 0$. **d**, Predicted fraction of IIE bound by AA-P. **e**, Predicted σ^F -RNAPol concentration during sporulation assuming that AB is allosteric (continuous line) or not allosteric with both AA and σ^F binding with high affinity (dashed line). **f**, Predicted concentration of σ^F (lower curves) or σ^A (upper curves) RNA polymerase holoenzyme formed 90 min after septation, as a function of an increase in IIE (including IIE complexed with AA-P) relative to the other sporulation proteins after 120 min of protein expression, assuming that AB is allosteric (continuous lines) or not allosteric with both AA and σ^F binding with high affinity (dashed lines). **g**, Predicted preseptational σ^F -RNAPol concentration when the sporulation proteins increase in parallel, with parameters as given in Supplementary Tables 1 and 2 (continuous line) or with $k_{off}(\sigma^F\text{-RNAPol}) = 0.02\text{ s}^{-1}$ (dashed line). On the horizontal axis, the concentration of σ^F is representative of σ^F , AA, AB and IIE. **h**, Predicted concentration of σ^F -RNAPol in a deterministic simulation (grey line) and in stochastic simulations (black lines) of the model.

line). However, if the increase in IIE is accompanied by an equimolar increase in its substrate, AA-P, almost all the σ^F is rapidly released (Fig. 4c, solid line); an increase in AA-P alone is not sufficient (Fig. 4c, dotted line). For the enzyme/substrate ratios found in the sporulating cell^{4,12,15} and the measured phosphatase rate and IIE-AA-P affinity (Supplementary Tables 1 and 2) the model predicts that enzyme and substrate will indeed appear together at the face of the asymmetric septum as the septum is being formed (Fig. 4d). The predicted increases in concentration of these two proteins in the smaller prespore are sufficient to trigger compartment-specific release of σ^F ; an increase in IIE phosphatase rate, for example by the removal of an inhibitor, is not required.

The model can also be applied to the competition of σ^F and σ^A for core RNA polymerase. It predicts the rapid formation of micromolar concentrations of σ^F holoenzyme (Fig. 4e) whenever the IIE concentration increases at least threefold over that in the predivisional cell (Fig. 4f). However, σ^F holoenzyme does not wholly supplant σ^A holoenzyme. The simulation mimics the known situation in the sporulating cell, in which both σ^A holoenzyme and σ^F holoenzyme are active¹⁶. If core RNA polymerase had the same affinity for σ^F as it has for σ^A , formation of σ^F holoenzyme would not require a septation-dependent increase in IIE activity (Fig. 4g). There would then be a risk that σ^F -specific transcription could occur in the predivisional cell once the concentrations of the sporulation proteins had increased sufficiently to permit the formation of micromolar concentrations of σ^F holoenzyme but before they had triggered septation.

The exceptional sensitivity of the model to a small increase in the concentration of AA derives from AB allostery (Fig. 4e, f). AA binding stabilizes the AB conformation that binds σ^F with low affinity; this enables the release of σ^F and prevents σ^F from rebinding once a critical AA concentration has been reached (Supplementary Information and Supplementary Fig. S10). Allostery also reduces the consumption of ATP by favouring the formation of AB-ADP-AA complexes, which reduces cycles of AA phosphorylation and dephosphorylation (Supplementary Information and Supplementary Fig. S11).

The model reveals several further, sophisticated features of the σ^F system that optimize its ability to regulate the initiation of sporulation. For example, several mechanisms are employed to control the ratios of σ^F and its regulatory proteins as well as the timing of septation; as a result the system is highly robust to stochastic fluctuations in gene expression that might arise because of low copy numbers¹⁷ (Fig. 4h; Supplementary Information and Supplementary Fig. S12), but it remains very sensitive to small changes in the concentrations of IIE and AA-P. Moreover, spontaneous initiation of sporulation is prevented by keeping the concentrations of σ^F and its regulators low in the vegetative cell. A detailed account of these features of the model will be published elsewhere, as will its predictions about the role of the transient genetic asymmetry^{18,19}, AB degradation²⁰ and preseptational AB-ADP-AA complex formation^{21,22}.

The value of any mathematical model depends critically on the accuracy of the parameters used. This study has benefited greatly from our knowledge of intracellular concentrations of the relevant components^{4,12,15} and from the fact that all the relevant kinetic constants are available from investigations of well-established *in vitro* systems². The resulting model accounts for numerous published results and correctly predicts the phenotype of essentially all mutants for which the biochemical effect of the mutation is known (Supplementary Information and Supplementary Table 3). In addition it has succeeded in predicting two novel features, AB allostery and a very low phosphatase rate *in vivo*, both of which prove to be essential components of the mechanisms that establish differential gene expression in this widely studied model system.

METHODS

Scatchard plot. A Scatchard plot was generated from AB-AA binding data obtained by SPR analysis. For the SPR measurements, concentrations of free AA in solution were determined with the use of AB as a reporter immobilized on the sensor chip. A standard curve of AA concentration against response units was produced with known concentrations of AA. The amount of free AA in AB-AA mixtures was then determined by comparison with this standard curve. The amount of bound AA was calculated from the total concentration of AA in the mixture. For this analysis the non-phosphorylatable mutant AA-S58A was used. Measurements with wild-type AA were also attempted. However, we were unable to obtain reliable data at the very high AB/AA ratios required for some of the measurements because it was not possible to eliminate phosphorylation of the AA by traces of ATP in the mixture.

SPR methods have been described previously¹⁵. AA–AB mixtures were generated with 4 μ M AA and 0.1–20 μ M AB. Because of a systematic error in the determination of protein concentrations the results have been scaled by a factor of 1.3 so that $\nu \leq 2$ in Fig. 2.

Fluorescence spectroscopy and enzyme kinetics. Fluorescence spectroscopy and measurements of enzyme kinetics were performed as described previously^{6–8}. Some of the results in Fig. 3 are derived from previously published data^{6,8}.

Numerical methods. The set of stiff ordinary differential equations (ODE) was solved by employing suitable Matlab ODE solvers. The stochastic simulation was performed with Matlab Simbiology using the stochastic simulation algorithm SSA, which is equivalent to solving the chemical Master equation for the system.

Received 19 January; accepted 21 February 2006.

- Hilbert, D. W. & Piggot, P. J. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol. Mol. Biol. Rev.* **68**, 234–262 (2004).
- Yudkin, M. D. & Clarkson, J. Differential gene expression in genetically identical sister cells: the initiation of sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **56**, 578–589 (2005).
- Duncan, L., Alper, S., Arigoni, F., Losick, R. & Stragier, P. Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. *Science* **270**, 641–644 (1995).
- Lord, M., Barillà, D. & Yudkin, M. D. Replacement of vegetative σ^A by sporulation-specific σ^F as a component of the RNA polymerase holoenzyme in sporulating *Bacillus subtilis*. *J. Bacteriol.* **181**, 2346–2350 (1999).
- Diederich, B. et al. Role of interactions between SpoIIAA and SpoIIAB in regulating cell-specific transcription factor σ^F of *Bacillus subtilis*. *Genes Dev.* **8**, 2653–2663 (1994).
- Clarkson, J., Shu, J.-C., Harris, D. A., Campbell, I. D. & Yudkin, M. D. Fluorescence and kinetic analysis of the SpoIIAB phosphorylation reaction, a key regulator of sporulation in *Bacillus subtilis*. *Biochemistry* **43**, 3120–3128 (2004).
- Clarkson, J., Campbell, I. D. & Yudkin, M. D. Physical evidence for the induced release of the *Bacillus subtilis* transcription factor, σ^F , from its inhibitory complex. *J. Mol. Biol.* **340**, 203–209 (2004).
- Clarkson, J., Campbell, I. D. & Yudkin, M. D. Efficient regulation of σ^F , the first sporulation-specific sigma factor in *B. subtilis*. *J. Mol. Biol.* **342**, 1187–1195 (2004).
- Shu, J.-C., Clarkson, J. & Yudkin, M. D. Studies of SpoIIAB mutant proteins elucidate the mechanisms that regulate the developmental transcription factor σ^F in *Bacillus subtilis*. *Biochem. J.* **384**, 169–178 (2004).
- Garsin, D. A., Duncan, L., Paskowitz, D. M. & Losick, R. The kinase activity of the antisigma factor SpoIIAB is required for activation as well as inhibition of transcription factor σ^F during sporulation in *Bacillus subtilis*. *J. Mol. Biol.* **284**, 569–578 (1998).
- Lee, C.-S., Lucet, I. & Yudkin, M. D. Fate of the SpoIIAB*–ADP liberated after SpoIIAB phosphorylates SpoIIAA of *Bacillus subtilis*. *J. Bacteriol.* **182**, 6250–6253 (2000).
- Lucet, I., Borriss, R. & Yudkin, M. D. Purification, kinetic properties, and intracellular concentration of SpoII_E, an integral membrane protein that regulates sporulation in *Bacillus subtilis*. *J. Bacteriol.* **181**, 3242–3245 (1999).
- Schroeter, R., Schlisio, S., Lucet, I., Yudkin, M. D. & Borriss, R. The *Bacillus subtilis* regulator protein SpoII_E shares functional and structural similarities with eukaryotic protein phosphatases 2C. *FEMS Microbiol. Lett.* **174**, 117–123 (1999).
- Eisenstadt, E., Fisher, S., Der, C. L. & Silver, S. Manganese transport in *Bacillus subtilis* W23 during growth and sporulation. *J. Bacteriol.* **113**, 1363–1372 (1973).
- Magnin, T., Lord, M. & Yudkin, M. D. Contribution of partner switching and SpoIIAA cycling to regulation of σ^F activity in sporulating *Bacillus subtilis*. *J. Bacteriol.* **179**, 3922–3927 (1997).
- Fujita, M. & Sadaie, Y. Rapid isolation of RNA polymerase from sporulating cells of *Bacillus subtilis*. *Gene* **221**, 185–190 (1998).
- Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
- Frandsen, N., Barák, I., Karmazyn-Campelli, C. & Stragier, P. Transient gene asymmetry during sporulation and establishment of cell specificity in *Bacillus subtilis*. *Genes Dev.* **13**, 394–399 (1999).
- Dworkin, J. & Losick, R. Differential gene expression governed by chromosomal spatial asymmetry. *Cell* **107**, 339–346 (2001).
- Pan, Q., Garsin, D. A. & Losick, R. Self-reinforcing activation of a cell-specific transcription factor by proteolysis of an anti-sigma factor in *Bacillus subtilis*. *Mol. Cell* **8**, 873–883 (2001).
- Carniol, K., Eichenberger, P. & Losick, R. A threshold mechanism governing activation of the developmental regulatory protein σ^F in *Bacillus subtilis*. *J. Biol. Chem.* **279**, 14860–14870 (2004).
- King, N., Dreesen, O., Stragier, P., Pogliano, K. & Losick, R. Septation, dephosphorylation, and the activation of σ^F during sporulation in *Bacillus subtilis*. *Genes Dev.* **13**, 1156–1167 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J.-C. Shu for providing unpublished SPR and AA phosphorylation data. This work was supported by the BBSRC UK. D.I. is a Junior Research Fellow at St John's College, Oxford, and is supported by a EPSRC DTA studentship. I.D.C. acknowledges financial support from the Wellcome Trust and the NIH-funded Cell Migration Consortium.

Author Contributions D.I. developed the model and conducted the simulations. J.C. performed the experiments and all the authors contributed to the concepts and writing of the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.I. (iber@maths.ox.ac.uk) or M.D.Y. (michael.yudkin@kellogg.ox.ac.uk).

Cleavage of pre-tRNAs by the splicing endonuclease requires a composite active site

Christopher R. Trotta¹, Sergey V. Paushkin¹, Meenal Patel¹, Hong Li² & Stuart W. Peltz¹

Splicing is required for the removal of introns from a subset of transfer RNAs in all eukaryotic organisms. The first step of splicing, intron recognition and cleavage, is performed by the tRNA-splicing endonuclease, a tetrameric enzyme composed of the protein subunits Sen54, Sen2, Sen34 and Sen15. It has previously been demonstrated that the active sites for cleavage at the 5' and 3' splice sites of precursor tRNA are contained within Sen2 and Sen34, respectively^{1–3}. A recent structure of an archaeal endonuclease complexed with a bulge–helix–bulge RNA has led to the unexpected hypothesis that catalysis requires a critical 'cation- π sandwich' composed of two arginine residues that serve to position the RNA substrate within the active site⁴. This motif is derived from a cross-subunit interaction between the two catalytic subunits. Here we test the role of this interaction within the eukaryotic endonuclease and show that catalysis at the 5' splice site requires the conserved cation- π sandwich derived from the Sen34 subunit in addition to the catalytic triad of Sen2. The catalysis of pre-tRNA by the eukaryotic tRNA-splicing endonuclease therefore requires a previously unrecognized composite active site.

Removal of introns from tRNA molecules in the archaeal and eukaryotic kingdoms requires a tRNA-splicing endonuclease to cleave the 5' and 3' splice sites. The combination of structural and biochemical studies of the archaeal and yeast tRNA-splicing endonuclease has led to a clear understanding of how all tRNA endonucleases arrange two distinct active sites to achieve cleavage at the 5' and 3' splice sites of an intron^{5–12}. Further analysis of a co-crystal consisting of the archaeal endonuclease complexed with a bulge–helix–bulge (BHB) tRNA substrate has led to the identification of key

amino-acid residues involved in RNA cleavage by the archaeal endonuclease⁴. The active site of the archaeal endonuclease of *Archaeoglobus fulgidus* is composed of a catalytic triad of residues, Tyr 246, His 257 and Lys 287, such that the scissile phosphate of precursor tRNA is situated at the centre that is formed by these amino acids. The importance of the conserved histidine residue has been demonstrated through mutagenesis in both archaeal and eukaryotic tRNA-splicing endonucleases^{5,8}. The co-crystal also revealed the presence of an arginine residue (Arg 280) in close proximity, about 4 Å, to each of the active sites. This arginine residue, together with a second (Arg 302), forms a 'cation- π sandwich' that seems to be involved in the recognition and stabilization of the interaction between the endonuclease and the tRNA substrate, allowing subsequent catalysis⁴. The cation- π sandwich functions like a pair of tweezers to pinch a flipped-out purine base. In particular, Arg 280 is positioned in a configuration that allows its guanidinium group to form a double hydrogen bond with a non-bridging phosphate oxygen following the last bulge nucleotide and a 2'-hydroxyl oxygen atom of the first bulge nucleotide of the three-nucleotide bulge required for cleavage by the archaeal endonuclease. This interaction facilitates a significantly bent bulge phosphate backbone that is critical for positioning the scissile phosphate for cleavage. Arg 302 is not universally conserved and corresponds to a tryptophan residue in both Sen2 (Trp 348) and Sen34 (Trp 271) of the yeast endonuclease, whereas Arg 280 is universally conserved in all archaeal and eukaryotic tRNA-splicing endonucleases (Arg 321 of Sen2 and Arg 243 of Sen34 of yeast endonuclease). Mutation of Arg 280 renders the archaeal endonuclease inactive at both splice sites⁴. The cation- π sandwich thought to be required for the 5' splice site is derived from the catalytic subunit situated at the 3' splice site. Conversely, the cation- π sandwich found at the 3' splice site is derived from the catalytic subunit responsible for the 5' splice site cleavage.

At present it is difficult to ascertain the significance of the cross-subunit interdependence for cleavage of RNA substrate within the

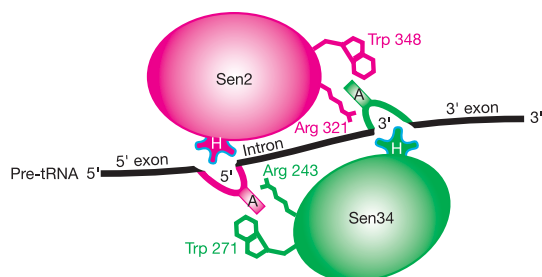


Figure 1 | Hypothetical model of the cation- π sandwich of the eukaryotic tRNA-splicing endonuclease. The active sites (H) of the eukaryotic tRNA-splicing endonuclease reside within the Sen2 and Sen34 subunits. Cleavage at the 5' splice site is proposed to require the active site of Sen2 and the cation- π sandwich of Sen34 (consisting of arginine (Arg) 243 and tryptophan (Trp) 271), and cleavage at the 3' splice would require the active site of Sen34 and the cation- π sandwich of Sen2 (consisting of Arg 321 and Trp 348). Pre-tRNA, 5' and 3' splice sites are designated. A flipped-out adenine base (A) at the 5' and 3' splice sites is shown stacked by the cation- π sandwich.

Table 1 | Inhibition of cleavage at either the 5' or 3' splice site

Mutant complex subunit	Percentage cleavage inhibition (splice site)	
	Pre-tRNA ^{Phe}	Pre-tRNA ^{Trp}
Sen34 His217Ala	88 (3')	96 (3')
Sen34 Arg243Ala	81 (5')	94 (5')
Sen34 Trp271Ala	70 (5')	99 (5')
Sen34 His217Ala/Arg243Ala	65 (5' and 3')	81 (5' and 3')
Sen2 His297Ala	74 (5')	88 (5')
Sen2 Arg321Ala	No inhibition	No inhibition
Sen2 Trp348Ala	No inhibition	No inhibition
Sen2 His297Ala/Arg321Ala	74 (5')	88 (5')

¹PTC Therapeutics, 100 Corporate Court, South Plainfield, New Jersey 07080, USA. ²Department of Chemistry and Biochemistry, Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA.

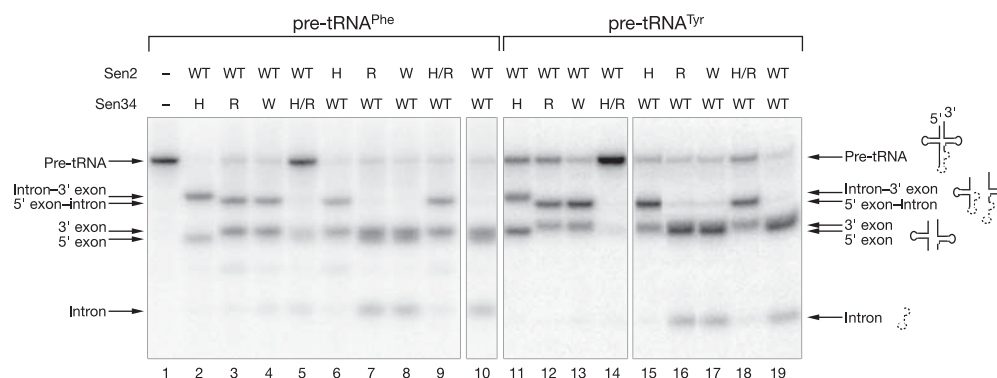


Figure 2 | Cleavage of yeast pre-tRNA^{Phe} and pre-tRNA^{Tyr} by mutant tRNA-splicing endonuclease. Enzymes containing mutations of the key amino acids in Sen2 and Sen34 were purified and tested in tRNA-splicing cleavage reactions (see Methods). Cleavage products are labelled from top to bottom as: pre-tRNA, intron-3' exon, 5' exon-intron, 3' exon, 5' exon and intron.

intron. Mutations to alanine are indicated by single-amino-acid designation: H, histidine; R, arginine; W, tryptophan; H/R, histidine/arginine double mutation. WT, wild type. A minor product is seen between exons and intron during cleavage with pre-tRNA^{Phe} but not with pre-tRNA^{Tyr}. This product is probably due to a non-specific nuclease that purifies with the enzyme complex.

archaeal endonuclease because the active sites are symmetrically disposed. This is not true of the eukaryotic enzyme. The eukaryotic tRNA-splicing endonuclease is composed of four independent subunits that have been proposed to be derived from several gene duplication events followed by subsequent specialization in which the subunits evolved unique roles in splicing of eukaryotic pre-tRNAs^{5,13,14}. Specifically, the Sen2 and Sen34 subunits have the highest degree of homology with the archaeal active site and a mutation of the critical histidine residue has shown that they provide the active site for the 5' and 3' splice site, respectively³. Despite the conserved nature of the active site, cleavage by the eukaryotic tRNA-splicing endonuclease is governed by a set of rules distinct from those of the archaeal tRNA endonuclease. Notably, recognition of eukaryotic precursor tRNA requires the mature domain for proper positioning of the 5' splice site and a conserved interaction known as the A-I (anticodon-intron) base pair that renders the 3' splice site competent for cleavage by the tRNA endonuclease^{6,15-17}.

On the basis of the conservation of the catalytic site and the amino-acid residues that form the archaeal cation- π sandwich, we proposed that cleavage of pre-tRNA by the eukaryotic tRNA-splicing endonuclease would be governed by a cation- π sandwich such that cleavage at the 5' splice site would require the cation- π sandwich of Sen34 and cleavage at the 3' splice site would require the cation- π sandwich of Sen2 (Fig. 1). To test this hypothesis we examined the effects of mutations in the amino-acid residues of the eukaryotic endonuclease corresponding to Arg 280 and Arg 302 of the cation- π sandwich of the archaeal endonuclease on cleavage of pre-tRNA. A series of mutants in either the 5' splice site subunit, Sen2, or the 3' splice site subunit, Sen34, were constructed by using quick-change mutagenesis and all four subunits of the enzyme were co-expressed in wild-type yeast, allowing the purification of heterologous yeast enzyme complexes (Supplementary Methods). Eight distinct mutant complexes were generated that included the following active-site pairs: Sen2 (wild-type (WT))/Sen34 (His217Ala), Sen2 (WT)/Sen34 (Arg243Ala), Sen2(WT)/Sen34 (Trp271Ala), Sen2 (WT)/Sen34 (His217Ala/Arg243Ala), Sen2 (His297Ala)/Sen34 (WT), Sen2 (Arg321Ala)/Sen34 (WT), Sen2 (Trp348Ala)/Sen34 (WT) and Sen2 (His297Ala/Arg321Ala)/Sen34 (WT) (see Supplementary Information). The activity of the purified enzyme complexes for the catalysis of pre-tRNA^{Phe} and pre-tRNA^{Tyr} substrates was evaluated (Supplementary Methods). As expected, mutation of the conserved active-site histidine residue in Sen2 (His297) or Sen34 (His217) impaired cleavage at the 5' splice site and the 3' splice site, respectively (Fig. 2, lanes 2, 6, 11 and 15). Inhibition of cleavage at the 3' splice site was 88% and 96% for Sen34 (His217Ala) on pre-tRNA^{Phe} and pre-tRNA^{Tyr}, respectively. Inhibition of cleavage at the 5' splice site was 74% and 88% for Sen2 His297Ala on pre-tRNA^{Phe}

and pre-tRNA^{Tyr}, respectively. Thus, the mutated enzyme complex contains one fully functional active site and one severely compromised active site that function independently during the cleavage of yeast pre-tRNA. Remarkably, mutation of either Arg 243 or Trp 271 of Sen34 significantly impaired cleavage at the 5' splice site by 81% and 70% on tRNA^{Phe} (Fig. 2, compare lanes 3 and 4 with lane 6) and 94% and 99% for pre-tRNA^{Tyr} (Fig. 2, compare lanes 12 and 13 with lane 15), respectively. The magnitude of deficiency for mutation of the cation- π sandwich of Sen34 is similar to that of mutation of the active-site histidine of Sen2 for 5' splice site cleavage (Table 1). Thus, the Sen34 subunit, which provides the active site for the 3' splice site cleavage, also has a crucial function in recognition and catalysis at the 5' splice site by the Sen2 active-site subunit. This is further demonstrated by the double mutant of Sen34 (His217Ala/Arg243Ala), in which cleavage is significantly impaired

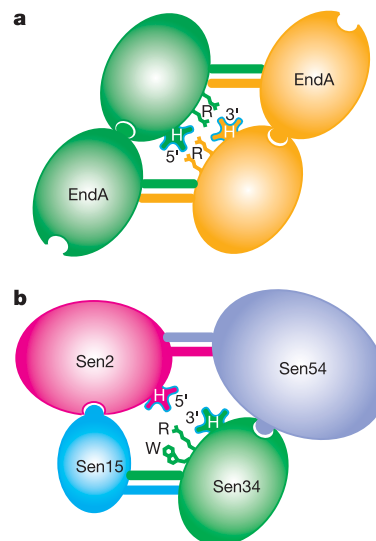


Figure 3 | Model of the archaeal and eukaryotic tRNA-splicing endonuclease. tRNA-splicing endonucleases from Archaea (a) and Eukarya (b) are configured as tetrameric enzymes with catalytic subunits contributing to cleavage at the 5' and 3' splice sites of tRNA substrates. The active site of the archaeal enzyme is a composite of a catalytic triad from one subunit (H) and a cation- π sandwich from the other subunit (R)⁴. Cleavage by the eukaryotic endonuclease requires the cation- π sandwich (R/W) of Sen34 subunit for cleavage only at the 5' splice site. Thus, the 5' splice site active site is a composite of catalytic triad from Sen2 (H) and an arginine and tryptophan cation- π sandwich of Sen34. Cleavage at the 3' splice site does not require the conserved cation- π sandwich of Sen2.

by more than 65% for pre-tRNA^{Phe} and by more than 81% in cleavage for pre-tRNA^{Tyr} at both the 5' and 3' splice sites (Fig. 2, lanes 5 and 14). The 5' splice site of eukaryotic tRNA substrates is usually found in an undefined loop and it has been proposed that cleavage of the 5' splice site requires both recognition of the mature domain of the tRNA precursor and rearrangement of the 5' splice site^{6,15,18,19}. Together with the observed arrangement of RNA within the active site of archaeal endonuclease, these results suggest that the cation- π sandwich provided by the Sen34 subunit of the eukaryotic endonuclease functions to precisely position the conserved arginine residue forming the basis for the structural rearrangement at the pre-tRNA 5' splice site.

The mutations tested did not always completely abolish cleavage for some pre-tRNA substrates (for example, Table 1 compare Sen34 His242Ala pre-tRNA^{Phe} to pre-tRNA^{Tyr}). We speculate that the residual activity seen with these mutants could indicate that, for some pre-tRNA substrates, recognition or cleavage by the endonuclease is not the rate-limiting step under these assay conditions. Further analysis of the impact of these mutations on the cleavage of different pre-tRNA substrates under various conditions will allow resolution of this observation.

We also tested the role of the conserved arginine and tryptophan residues in Sen2 for cleavage at the 3' splice site. In contrast to the requirement for the cation- π sandwich for cleavage of the 5' splice site in yeast pre-tRNA, the results demonstrated that mutation of Arg 321 or Trp 348 in Sen2 did not abolish cleavage at the 3' splice site (Fig. 2, lanes 7, 8, 16 and 17). Cleavage of the 3' splice site is independent of the cation- π sandwich of Sen2. We propose that cleavage at the 3' splice site in eukaryotic precursor tRNA evolved to depend on the conserved A-I base pair, which serves to orient the 3' splice site properly for cleavage^{16,17}. This observation is consistent with previous results showing that cleavage at the 3' splice requires this motif and is independent of recognition of the mature domain of pre-tRNA^{6,19}.

The results presented here provide evidence for the use of a composite active site formed by key amino-acid residues from distinct subunits for cleavage by an RNA endonuclease (Fig. 3). This strongly supports the unexpected finding of interdependence between the catalytic subunits of the archaeal tRNA-splicing endonuclease and deepens our understanding of how the endonuclease active site has evolved to accommodate different RNA structures found at the splice sites of various RNA substrates. The eukaryotic endonuclease has apparently retained the requirement for a composite active site in promoting catalysis of the 5' splice site of precursor tRNA by Sen2 (Fig. 3b), whereas cleavage at the 3' splice site has evolved to become independent of this function. Similar results obtained with two different yeast pre-tRNAs and the structural conservation of intron-containing tRNA in yeast¹⁸ strongly indicate a universal function of the cation- π sandwich for cleavage of the 5' splice site of pre-tRNA by the eukaryotic endonuclease. We propose that the catalytic subunits of the eukaryotic endonuclease have coevolved with tRNA intron motifs, similarly to the proposed evolution of the archaeal endonuclease¹². It has recently been suggested that the endonuclease subunits might have a function in additional RNA catalytic processes^{10,20}. Thus, in addition to the pre-tRNA substrate, evolution of the active sites might also be governed by their roles in processing or metabolizing other RNA substrates.

METHODS

Purification of yeast tRNA endonuclease. Yeast tRNA-splicing endonuclease was generated by expression of the appropriately mutated His/Flag-tagged Sen2 or His/Flag-tagged Sen34 along with Sen54 and Sen15 on galactose-inducible plasmid with GAL1/GAL10 promoters in wild-type yeast followed by subsequent purification from yeast cytoplasmic extract. Details of the endonuclease purification protocol are given in Supplementary Methods. Purified enzyme was quantified by SDS-PAGE and an equal amount of enzyme was used for tRNA-splicing reactions. **Mutagenesis of Sen2 and Sen34.** Mutants in yeast Sen2 and Sen34 were created by quick-change mutagenesis in accordance with the manufacturer's protocol (Stratagene). Oligonucleotide sequences are given in the Supplementary

Methods. All mutations were confirmed by sequencing (Seqwright).

tRNA-splicing endonuclease assays. tRNA-splicing endonuclease reactions were performed as described previously²⁰. Pre-tRNA^{Phe} and pre-tRNA^{Tyr} were generated by transcription with T7 RNA polymerase and labelled with [α -³²P]UTP as described previously^{21,22}. A detailed description of cleavage analysis is given in Supplementary Methods.

Received 9 December 2005; accepted 23 March 2006.

- Abelson, J., Trotta, C. R. & Li, H. tRNA splicing. *J. Biol. Chem.* **273**, 12685–12688 (1998).
- Ho, C. K., Rauhut, R., Vijayraghavan, U. & Abelson, J. Accumulation of pre-tRNA splicing '2/3' intermediates in a *Saccharomyces cerevisiae* mutant. *EMBO J.* **9**, 1245–1252 (1990).
- Trotta, C. R. *et al.* The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* **89**, 849–858 (1997).
- Xue, S., Calvin, K. & Li, H. Structural mechanism of RNA recognition and cleavage by a splicing endonuclease. *Science* doi:10.1126/science.1126629 (in the press).
- Belfort, M. & Weiner, A. Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes. *Cell* **89**, 1003–1006 (1997).
- Di Nicola, N. E. *et al.* The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. *Cell* **89**, 859–866 (1997).
- Kleman-Leyer, K., Armbruster, D. W. & Daniels, C. J. Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell* **89**, 839–847 (1997).
- Li, H., Trotta, C. R. & Abelson, J. Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* **280**, 279–284 (1998).
- Li, H. & Abelson, J. Crystal structure of a dimeric archaeal splicing endonuclease. *J. Mol. Biol.* **302**, 639–648 (2000).
- Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. *Cell* **117**, 311–321 (2004).
- Randau, L. *et al.* The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge–helix–bulge motifs of joined tRNA halves. *Proc. Natl Acad. Sci. USA* **102**, 17934–17939 (2005).
- Tocchini-Valentini, G. D., Fruscoloni, P. & Tocchini-Valentini, G. P. Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc. Natl Acad. Sci. USA* **102**, 15418–15422 (2005).
- Bujnicki, J. M. & Rychlewski, L. Prediction of a common fold for all four subunits of the yeast tRNA splicing endonuclease: implications for the evolution of the EndA/Sen family. *FEBS Lett.* **486**, 328–329 (2000).
- Lykke-Andersen, J. & Garrett, R. A. RNA-protein interactions of an archaeal homotetrameric splicing endonuclease with an exceptional evolutionary history. *EMBO J.* **16**, 6290–6300 (1997).
- Reyes, V. M. & Abelson, J. Substrate recognition and splice site determination in yeast tRNA splicing. *Cell* **55**, 719–730 (1988).
- Abelson, J. Recognition of tRNA precursors: a role for the intron. *Science* **255**, 1390 (1992).
- Baldi, M. I., Mattoccia, E., Bufardecchi, E., Fabbri, S. & Tocchini-Valentini, G. P. Participation of the intron in the reaction catalyzed by the *Xenopus* tRNA splicing endonuclease. *Science* **255**, 1404–1408 (1992).
- Lee, M. C. & Knapp, G. Transfer RNA splicing in *Saccharomyces cerevisiae*. Secondary and tertiary structures of the substrates. *J. Biol. Chem.* **260**, 3108–3115 (1985).
- Fruscoloni, P., Baldi, M. I. & Tocchini-Valentini, G. P. Cleavage of non-tRNA substrates by eukaryal tRNA splicing endonucleases. *EMBO Rep.* **2**, 217–221 (2001).
- Volta, V. *et al.* Sen34p depletion blocks tRNA splicing *in vivo* and delays rRNA processing. *Biochem. Biophys. Res. Commun.* **337**, 89–94 (2005).
- Rauhut, R., Green, P. R. & Abelson, J. Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme. *J. Biol. Chem.* **265**, 18180–18184 (1990).
- Reyes, V. M. & Abelson, J. A synthetic substrate for tRNA splicing. *Anal. Biochem.* **166**, 90–106 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Romfo, A. Jacobson, J. Colacino and D. Glodowski for discussions and critical reading of the manuscript.

Author contributions C.R.T. and S.V.P. contributed equally to this work and together with M.P. performed the mutagenesis, purification and testing of endonuclease complexes. C.R.T., H.L., S.V.P. and S.W.P. wrote the manuscript together.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.R.T. (ctrotta@ptcbio.com).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Like crime, scientific misconduct ultimately doesn't pay. For the two high-profile cases of data fabrication in recent years — Jan Hendrik Schön of Bell Laboratories and Woo Suk Hwang of Seoul National University — the price was a ruined reputation and dismissal. Others found guilty of similar misdeeds have found themselves barred from doing research or from receiving federal funds (see *Nature Med.* **12**, 490–494; May 2006).

But what of the innocent parties caught up in the crossfire? Postdocs and graduate students working under a principal investigator who is found guilty of misconduct can find themselves tainted by association. At the very least, they will face setbacks in their careers — what use to any author is a retracted paper? And there is little point in asking your adviser for a letter of reference if he or she has been discredited.

People working under an adviser whom they know to be engaged in misconduct face a difficult choice. If they do nothing, then they stand a chance of completing their studies and moving on to a different lab — but if their adviser is eventually found out, they risk being judged as guilty by association. If, on the other hand, they choose to blow the whistle, life can become very uncomfortable. In many instances, young scientists who report incidences of misconduct by their seniors can face as much scrutiny as the people they have accused (see *Nature* **441**, 122–123; 2006).

To resolve these issues, and to minimize any risk of fraudulent behaviour, we need a new system that protects whistleblowers. Such a system would give universities, funding agencies and journals clear roles and guidelines for policing fraud. It would also, by necessity, shift the burden of proof from the accuser on to the accused.

If trainees find themselves in the unfortunate position of having to point out questionable research practices, they should be able to do so without worrying that they could damage their career. This is especially true when you consider that not blowing the whistle could have even worse long-term effects.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik

European Head Office, London
The Macmillan Building,
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)

Natureevents:
Matt Betteridge (4015)

UK/RoW/Ireland:
Niels Moeller (4953)

Scandinavia/Spain/Portugal:

Claudia Paulsen Young (4973)

France/Switzerland/Belgium/Italy:
Maria Ceraolo (4994)

Germany/Austria/The Netherlands:

Reya Silao (4970)

Business Development Manager:
Amelie Pequignot (4974)

Advertising Production Manager:

Stephen Russell

To send materials use London
address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Catherine Alexander

European Satellite Office

Germany: Patrick Phelan

Tel: +49 89 54 90 57 11

Fax: +49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

75 Varick Street,

9th Floor, New York,

NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,

2-37 Ichigayatamachi,

Shinjuku-ku,

Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

Asia-Pacific Sales Director:

Ayako Watanabe

e-mail: a.watanabe@natureasia.com

MOVERS

Mahendra Rao, vice-president, research, stem cells and regenerative medicine, Invitrogen, Carlsbad, California



2001-present: Associate professor of neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland

2001-05: Section chief, stem-cell division, neuroscience laboratory, National Institute of Aging, Washington DC

1999-present: Associate professor, National Centre for Biological Sciences, Bangalore, India

Poor eyesight quashed Mahendra Rao's early ambitions of becoming an air-force pilot, but his training as a medical officer at Bombay University steered his path to the United States — and a new vision of using stem cells in medicine.

While a resident in clinical medicine and neurology, Rao realized there were few treatments for his patients, so he turned to research. He left India for a PhD in neurobiology at the California Institute of Technology in Pasadena. Although a difficult career decision, given the change in lifestyle and income, it gave him the chance to work with leading biologists such as Paul Patterson. In Patterson's lab, Rao isolated neural-crest stem cells, the cells that regulate development of the peripheral nervous system.

Patterson also helped Rao forge future collaborations, notably conducting his postdoc at Case Western Reserve University in Cleveland, Ohio, with Story Landis, now director of the National Institute of Neurological Disorders and Stroke. There, Rao focused on factors affecting the fate of stem cells. Finding opportunities to exploit new technology was also part of Rao's strategy. Rao joined the University of Utah's school of medicine as an assistant professor, in part because it housed the first centre to generate transgenic mice using embryonic stem cells — a logical fit with his work demonstrating the relationships between the stages of stem-cell development.

After five years at Utah, Rao became an adjunct faculty member at the National Centre for Biological Sciences in Bangalore, India, where he continues to teach courses. It was his 2001 move to the US National Institute of Aging, however, that paired his stem-cell research with studies of mortality. The signalling mechanisms that regulate stem cells' ability to escape death change during ageing and age-related neurodegenerative disorders.

But federal restrictions on the use of stem cells slowed his work and eventually triggered a move to Invitrogen, where he will have freedom to pursue his research interests. "It's unlikely that I would have gone to industry without the current policy decisions," says Rao. "But, in California, I saw people who wanted to make it work rather than worrying about breaking the law."

Rao remains hopeful that one day the US federal government will support stem-cell research. Until then, he hopes others won't be deterred from pursuing their own interests in that area. "Do what you love, because it's what you will do best," he says.

Virginia Gewin

SCIENTISTS & SOCIETIES

International view from Japan

Young scientists need to communicate with their peers worldwide, but there is little opportunity to do this in Asia.

Most international scientific meetings are held in the United States or Europe, so the long distance and the expense, as well as difficulty communicating in English, make it hard for young Asian scientists to take part.

To deal with this problem, a group of graduate students in Kyoto brought some young Western scientists to Kyoto University's Graduate School of Biostudies and the Institute for Virus Research in March. We organized a seminar to provide an opportunity for young scientists from all over the world to exchange scientific experiences and form friendships and, perhaps, collaborations with each other.

We invited 15 graduate students and five postdoctoral fellows from abroad. All were the same generation as us, and had similar interests in fashion, sports and career goals, so we could enjoy talking with them in addition to exchanging scientific experience.

The first half of the four-day symposium focused on scientific presentations. We showcased 40 oral presentations and 110 posters covering a range of life-science topics.

In the second half, we held a group discussion at Ninna-ji temple in Kyoto,

focusing on similarities and differences in science, culture and lifestyle between Japan and other nations. Common concerns included balancing scientific and personal life, but cultural differences led to different approaches to dealing with this. Many Western participants said they came to the lab very early in order to have free time after supper, whereas many Japanese scientists tend to work very late.

The scientific ideal was the same for all of us: a love of science, an interest in problem-solving and a desire to improve communication with society as well as with colleagues.

This seminar was organized as part of the 21st Century Center of Excellence Program at Kyoto University, funded by the Japanese government. The focus on student presentations and collaborative relationships provided a valuable experience for both Japanese and foreign students. The seminar let us form collaborations and discuss the future of many scientific fields from an international perspective. We hope that international communication among young scientists will increase through more such seminars.

Keiko Muraki is a graduate student at the University of Kyoto, Japan.

▶ www.lif.kyoto-u.ac.jp/coe4th

GRADUATE JOURNAL

The many legs of fear

This week, at various intervals, I have become withdrawn, irrational and emotional. I have screamed as in a horror movie. No, not because my PCR didn't work. Because of wasps in the lab.

It may seem surprising to some, but despite doing a degree in biology, I am petrified of insects. As an undergraduate I remember getting my twin sister (who studied English) to tape over the pictures of insects in my textbooks. Open windows in summer exams scared me more than the questions did.

Having tried very hard, I have learnt to hide my fear. In lectures with *Drosophila* pictures I just dip my head, or lower my glasses so I can't focus. Walking down the fly-geneticists' corridor in the zoology department I make sure not to catch a glimpse of the posters. It's the eyes and legs that really scare me.

This week in Oxford the sky is blue, the grass is green and my colleagues have discovered my fear. In reaction to this they did not agree to close all the windows. Nor did they swear to swat all the wee beasties. Instead they put a picture of a wasp on my computer desktop, and then sent me a photo titled 'Interesting fern mutant' showing the magnified head of a fly.

So this week I have learnt to try to overcome my fears, to raise my problems with my lab colleagues, and to persevere with something I don't like — all skills that would be handy in a PhD. Now I am planning my revenge...

Mhairi Dupré is a first-year PhD student in evolutionary developmental biology at the University of Oxford, UK.

The visible men

Or, down the multiversal rabbit hole.

Michael Moorcock

"That a cat's cradle?" Miss Brunner peered down at a naked Jerry Cornelius tangling his hands in a mess of guitar strings. A red Rickenbacker twelve lay beside him.

"It's twine theory, he said." Frank was absorbed in his own calculations covering the large slate propped on his mum's kitchen table. "He got a bit confused. Too many Es. Too much reverb." He followed her gaze. "G? Somewhere in the seventh dimension."

"He's a simple soul at heart. Easily led..." Major Nye stroked his pale moustache. He'd come in with Miss Brunner hoping to take Mrs Cornelius out. "Is she here at all?"

"Pictures with Colonel Pyat." Frank spoke spitefully. "IT at the Electric. I'll tell her you called." His horrible feet in a bowl of soapy water, he frowned over his equations. What had been in that third syringe?

"Pip," said Jerry. "Pip. Pip." The strings coiled into a neat pile and vanished. He beamed.

Frank wondered why Jerry could charm and he couldn't?

Jerry strolled into the basement room sniffing. At the window, Jerry stopped to test the bars. In the kitchen Jerry cursed as he felt about in the toaster. From the front door upstairs Jerry called through the letter box. They were all naked, save for black car-coats. Jerry stood up pulling on his underpants. "Sorry I'm not decent."

Miss Brunner turned away with a strangled word. "What...?"

"Interdimensional travel." Jerry knotted his wide tie, copping Frank's calculations. "Though not very sophisticated." He reached to rub out a figure.

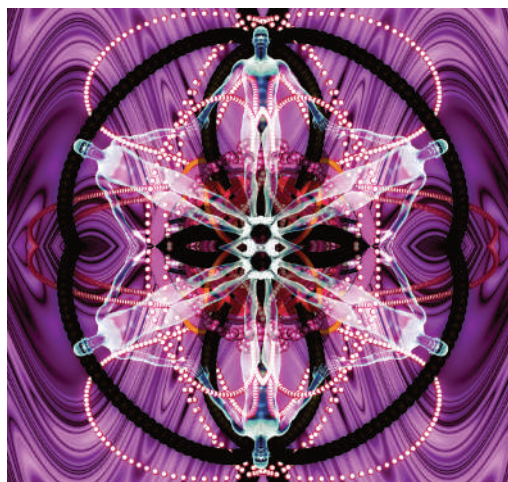
Pettishly, Frank slapped him. "Just the air cooling. Entropy factor. Anyway, your sides are all slightly different."

"All?" Jerry frowned at the versions of himself. "If I had a black hole they'd follow me into it. As it is..."

Frank scowled. "You and your bloody multiverse. Energy's bound to thin out if you're that profligate."

"Crap." Jerry holstered his vibragun. "Effectively energy's limitless. It's Mandelbrot, Frank. Each set's invisibly smaller. Or invisibly bigger. Depending where you start. You don't go through the multiverse — you go up and down scales of almost infinite but tiny variability. Only the mass

varies enormously, making them invisible. That's why we're all essentially the same." With scarcely any echo, identical voices came from each identical mouth: "Only after travelling through billions of sets do you start spotting major differences. The quasi-infinite, Frank. Think how many billions of multiversal planes of the Universe there are! Vast as it is, with my box you can step from one end to the other in about ten minutes. Go all the way round. Your mass compresses or expands accordingly. Once I realized space is a dimension of time, the rest was easy!"



"Pervert! You and your proliferating clones."

"Clones?" Miss Brunner licked her lips. "Are they edible?" She adjusted her powder-blue two-piece.

"They're not clones, they're versions. When you dash about the multiverse, this sort of thing happens. I prefer to shrink. But denser, you rip holes; drag things in. Nobody sees the universe next door because it's too big or too small. Fractional, of course, in multiversal terms. Problem is, bits of one universe get sucked into another. They're all so close. Déjà vu...?"

"Carry on like this, young man," Major Nye straightened his cap, "and you'll cause the end of matter. You'll have your chaos, all right!" Feelings hurt, he made for the basement door.

"That's ridiculous." Miss Brunner repaired her face. "Why aren't your clones..."

"Duplicates."

"Why aren't they too big or too small to see?"

"That's the whole trick." Jerry preened. Now in sync, his rippling duplicates fol-

lowed his every move. "Getting us all to the same scale. Expansion and compression. Your atoms only change mass, maintaining identity. See, we're either too huge to perceive the next universe or we're so massively tiny we merely pass through it without noticing it. Either way you can't see 'em. Until I use this little gadget."

With a disapproving pout, she clicked across the parquet.

"You change your mass relative to theirs, or vice versa, and they become visible. At first you feel a bit queasy, but you get used to it." Picking up the small black box from the table, he showed her the display, the triggers. "Have a go. It's easy. Everything's digitalized."

"Certainly not. I have enough trouble controlling my own world."

"But this gives you millions of alternatives. Immortality of sorts. Admittedly, the nearest billion or so are boringly alike. But most people, like you, love repetition..."

"Rot! Utter dissipation! Double Deutsch, I call it!" Grumpily, Major Nye closed the door. Through the bars they saw him climb area steps, pushing aside three more Jerrys staring at one another in some confusion.

Upstairs the front door opened.

"Oh, blimey!" Dismayed, Jerry peered around for a hiding place.

"Mum's back early."

"You'll have some explaining to do." Frank smirked.

But Jerry was already fiddling with his box and wires. As Mrs Cornelius waddled into the room, exuding a delicious smell of greasy fish, Jerry shrank into a corner, his duplicates following. Everyone stared after him.

"Fairlyland again!" Miss Brunner was contemptuous.

"The Major said Jerry 'ad a message. Where's 'e gorn?" Mrs Cornelius lifted huge blue suspicious eyes. A plump hand carried chips from her newspaper to her mouth.

"Climbing the bloody beanstalk, as usual." Defeated, Frank faded.

Mrs C roared.

Michael Moorcock claims to have invented the multiverse in 1961 but then he claims to have created London in 1965, the same year he created Jerry Cornelius, whose adventures won the Guardian Fiction Prize in 1977 despite the author's insistence of their authenticity. Colonel Pyat's final memoir *The Vengeance of Rome* appeared in January 2006. Moorcock is currently resting in a Texas institution.

JACEY